

**VENTAJAS Y DESVENTAJAS DEL ANÁLISIS DE UN DISEÑO DE MEDIDAS  
REPETIDAS MEDIANTE ANÁLISIS DE VARIANZA UNIVARIADO, MULTIVARIADO  
Y MODELOS LINEALES MIXTOS**

**ANDRÉS MAURICIO BERNAL PÉREZ  
JHON ALEXANDER ROZO FORERO**

**Trabajo de grado como requisito para optar al título de profesional en  
Matemáticas con Énfasis en Estadística**

**Directora  
YURI MARCELA GARCIA SAAVEDRA  
Magister en Estadística**

**UNIVERSIDAD DEL TOLIMA  
FACULTAD DE CIENCIAS  
MATEMÁTICAS CON ÉNFASIS EN ESTADÍSTICA  
IBAGUÉ – TOLIMA**

**2017**



UNIVERSIDAD DEL TOLIMA

FACULTAD DE CIENCIAS

PROGRAMA DE MATEMÁTICAS CON ÉNFASIS EN ESTADÍSTICA

ACTA DE SUSTENTACIÓN TRABAJO DE GRADO

**TÍTULO:** VENTAJAS Y DESVENTAJAS DEL ANÁLISIS DE VARIANZA UNIVARIADO, MULTIVARIADO Y MODELOS LINEALES MIXTOS

**AUTORES:** ANDRÉS MAURICIO BERNAL Cod. 070250342009  
JOHN ALEXANDER ROZO FORERO Cod. 070250102009

**DIRECTOR:** YURI MARCELA GARCÍA SAAVEDRA

**JURADOS:** ALFONSO SÁNCHEZ HERNÁNDEZ  
JULIE KIMBERLY RAMIREZ B.

**CALIFICACIÓN:** 4.7 (Meritorio)

☒ APROBÓ

☐ REPROBÓ

OBSERVACIONES:

FIRMAS

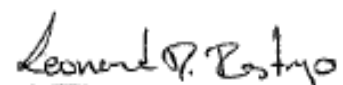
  
ALFONSO SÁNCHEZ HERNÁNDEZ

Jurado 1

  
YURI MARCELA GARCÍA SAAVEDRA  
Directora del Trabajo

  
JULIE KIMBERLY RAMIREZ B.

Jurado 2

  
LEONARDO DUVAN RESTREPO  
Director del Programa

Ciudad y fecha: Ibagué, 08 de Junio de 2017

## **DEDICATORIA**

A nuestros padres que nos enseñaron a no rendirnos y a seguir aprendiendo siempre, siendo nuestros grandes ejes de apoyo.

A nuestras novias Alejandra Manjarrez y Carolina Cerón quienes siempre están ahí brindándonos amor y ayudándonos siempre con mucho cariño.

A la Profesora Yuri Marcela García Saavedra quien con su gran trabajo, enseñanza y paciencia logramos superar este camino.

## **AGRADECIMIENTOS**

Primero que todo le damos gracias a Dios por acompañarnos siempre, guiarnos en todo nuestro proceso, por darnos salud e iluminarnos cada día.

Le damos gracias a nuestros padres ya que sin ellos nada de esto sería posible, por su apoyo incondicional y por brindarnos la oportunidad de salir a delante con tantos sacrificios, brindarnos amor y ser nuestros amigos en todos los momentos.

A nuestros hermanos, familiares y amigos que también han hecho posible este logro, con palabras de motivación, que nos proporcionaron más ganas de salir adelante.

A nuestras parejas que sin ellas no habría sido posible, porque su apoyo, su tiempo, ayuda constante su perseverancia, cariño, amor y comprensión, nos hacían creer más en lograr el objetivo.

A la profesora Yuri Marcela García, le agradecemos mucho, nos ha aportado inmensamente, por su tolerancia y comprensión, por su tiempo y conocimiento que nos transmitió a lo largo de este proceso.

## GLOSARIO

**ANÁLISIS DE LA VARIANZA:** es una técnica estadística que sirve para decidir/determinar si las diferencias que existen entre las medias de tres o más grupos (niveles de clasificación) son significativamente diferentes. Las técnicas de ANOVA se basan en la partición de la varianza para establecer si la varianza explicada por los grupos formados es suficientemente mayor que la varianza residual o no explicada.

**COEFICIENTE DE VARIACIÓN:** es la relación entre la desviación típica de una muestra y su media.

**CORRELACIÓN:** medida de la relación existente entre dos variables. Su valor está comprendido entre  $-1$  y  $1$ . Si es negativo la relación entre las variables es inversa, es decir, a medida que aumentan los valores de una decrecen los de la otra. Si es positivo la asociación es directa, es decir, los valores de una variable aumentan con la otra. Un valor de cero indica ausencia de relación.

**COVARIANZA:** representa la media del producto de las desviaciones de dos variables en relación a su media.

**DATOS ALEATORIOS:** son datos obtenidos al azar a partir de una población a los cuales no se les ha dado ninguna prioridad, es decir, todos tienen la misma probabilidad de ser elegidos.

**DESVIACIÓN TÍPICA:** valor mayor o igual a cero que mide la dispersión de una característica de los individuos alrededor de la media del grupo. Los sujetos serán más parecidos u homogéneos entre sí cuanto más próxima a cero esté la desviación típica. Su unidad de medida coincide con la unidad de medida de la variable original. Es la raíz cuadrada de la varianza.

**DISEÑO DE EXPERIMENTOS:** método estadístico cuyo objetivo es estudiar cómo cambian los valores de una variable respuesta cuando se modifican los valores de una o varias variables independientes, denominadas factores experimentales. Un experimento bien diseñado puede ser el punto de partida para establecer relaciones causales entre las variables estudiadas.

**ESTADÍSTICA:** ciencia que estudia los fenómenos aleatorios. Es un área de conocimiento específico de las Matemáticas que comenzó a desarrollarse a mediados del siglo XVII. Sus técnicas permiten resumir grandes cantidades de información, estudiar la relación entre variables, investigar la causa de algunos sucesos o predecir la evolución de un fenómeno en el tiempo y en el espacio, entre otras cosas.

**FACTOR:** variable que se incluye en un modelo con el propósito de explicar la variación en la variable respuesta. Véase variable independiente o variable explicativa.

**HETEROSCEDASTICIDAD:** hipótesis de no igualdad de varianzas poblacionales en distintos grupos.

**HIPÓTESIS:** cualquier teoría que formule posibles líneas de trabajo experimental.

**HIPÓTESIS NULA:** afirmación establecida por el investigador sobre la población de estudio cuando realiza un test o contraste de hipótesis. Esta hipótesis siempre se asume verdadera a menos que los datos de la muestra proporcionen evidencia de lo contrario. Se suele formular mediante una negación o una igualdad.

**HOMOSCEDASTICIDAD:** hipótesis de igualdad de varianzas poblacionales en distintos grupos.

**INDEPENDENCIA:** son datos que no están ligados entre sí.

**INTERVALO DE CONFIANZA:** rango de valores que, con una cierta confianza, contiene al parámetro poblacional que se pretende conocer. el intervalo de confianza se construye a partir de la información de la muestra y es una de las herramientas utilizadas para extrapolar los resultados a la población.

**MEDIDA REPETIDA:** método estadístico el cual obtiene sus datos a través de múltiples observaciones al sujeto que va a hacer parte del estudio.

**MODELO ESTADÍSTICO:** es una ecuación matemática que reproduce los fenómenos que observamos de la forma más exacta posible. Para ello tiene en cuenta los datos suministrados y la influencia que el azar tiene en estas observaciones.

**MODELO MIXTO:** es aquel modelo estadístico que involucra efectos fijos y efectos aleatorios.

**SIMETRÍA:** es una medida que refleja si los valores muestrales se extienden o no de igual forma a ambos lados de la media.

**VALOR P:** valor comprendido entre 0 y 1 que está asociado a un test de hipótesis. Es la probabilidad de encontrar un resultado como el obtenido en la muestra, o incluso más extraño, cuando la hipótesis nula es cierta. La hipótesis nula será rechazada siempre que esta probabilidad sea muy pequeña, inferior al nivel de significación preestablecido.

**VARIABLE:** cualquier característica de los individuos que componen la muestra que toma un valor diferente para cada uno de ellos. La edad y el sexo son variables habituales en muchas investigaciones.

**VARIANZA:** valor mayor o igual a cero que mide la dispersión de una característica de los individuos alrededor de la media del grupo. Los sujetos serán más parecidos u homogéneos entre sí cuanto más próxima a cero esté la varianza. Su unidad de medida

es la unidad de medida de la variable original elevada al cuadrado. Su valor coincide con el cuadrado de la desviación típica.



## CONTENIDO

Pág.

<b>INTRODUCCIÓN .....</b>	<b>15</b>
<b>1. ANTECEDENTES .....</b>	<b>16</b>
<b>2. JUSTIFICACIÓN .....</b>	<b>6</b>
<b>3. OBJETIVOS .....</b>	<b>7</b>
3.1 OBJETIVO GENERAL .....	7
3.2 OBJETIVOS ESPECÍFICOS .....	7
<b>4. METODOLOGÍA.....</b>	<b>8</b>
<b>5. MARCO TEÓRICO.....</b>	<b>9</b>
5.1 INTRODUCCIÓN A LAS MEDIDAS REPETIDAS.....	9
5.1.1 Conceptos básicos.....	11
5.2 ANÁLISIS DE VARIANZA UNIVARIADO (ANOVA) .....	16
5.2.1 Tabla ANOVA.....	18
5.2.2 Supuestos. ....	19
5.3 ANÁLISIS DE VARIANZA MULTIVARIADO (MANOVA).....	19
5.3.1 Contrastes de hipótesis.....	21
5.4 MEDIDAS REPETIDAS EN Q-MUESTRAS.....	23
5.5 MODELO DE DOS FACTORES DENTRO DE UN SUJETO Y UN FACTOR ENTRE SUJETOS.....	27
5.6 MODELOS MIXTOS.....	33
5.6.1. Ventajas.....	36
5.6.2. Estructura de coovarianza.....	40
5.7 CRITERIO DE INFORMACIÓN AKAIKE (AIC).....	43
5.8 CRITERIO DE INFORMACIÓN BAYESIANA (BIC) .....	44

<b>6. APLICACIONES.....</b>	<b>45</b>
<b>6.1 ASPECTOS GENERALES.....</b>	<b>45</b>
<b>6.2 Análisis De Varianza “ANOVA” .....</b>	<b>48</b>
<b>6.2.1 Prueba de Normalidad .....</b>	<b>49</b>
<b>6.2.2 Prueba de Homogeneidad .....</b>	<b>50</b>
<b>6.2.3 Prueba de Independencia .....</b>	<b>51</b>
<b>6.3 ANÁLISIS MULTIVARIADO “MANOVA” .....</b>	<b>52</b>
<b>6.3.1 Prueba de Normalidad. ....</b>	<b>53</b>
<b>6.3.2 Prueba de Homoscedasticidad.....</b>	<b>53</b>
<b>6.3.3 Prueba De Independencia.....</b>	<b>55</b>
<b>6.4 MODELOS LINEALES MIXTOS.....</b>	<b>55</b>
 <b>7. CONCLUSIONES .....</b>	 <b>63</b>
 <b>RECOMENDACIONES Y TRABAJOS FUTUROS.....</b>	 <b>65</b>
 <b>REFERENCIAS BIBLIOGRAFICAS.....</b>	 <b>66</b>
 <b>ANEXOS .....</b>	 <b>68</b>

## LISTA DE TABLAS

	<b>Pág.</b>
<b>Tabla 1:</b> Análisis de varianzas de 1 factor	18
<b>Tabla 2:</b> Medidas repetidas en q – grupos.....	24
<b>Tabla 3:</b> Factores dentro de sujetos (A y B) .....	28
<b>Tabla 4:</b> Estadística descriptiva por semana.....	47
<b>Tabla 5:</b> Estadística descriptiva por tratamiento.....	48
<b>Tabla 6:</b> ANOVA.....	48
<b>Tabla 7:</b> Prueba de normalidad Shapiro Wilk.....	50
<b>Tabla 8:</b> Prueba de Levene para Homoscedasticidad de varianza.....	51
<b>Tabla 9:</b> MANOVA.....	52
<b>Tabla 10:</b> Prueba de normalidad Shapiro Wilk.....	53
<b>Tabla 11:</b> Prueba de Levene.....	54
<b>Tabla 12:</b> AIC y BIC de los modelos escogidos.....	56
<b>Tabla 13:</b> Resultados del coeficiente de verosimilitud... ..	58
<b>Tabla 14:</b> Anova del modelo 6.....	59
<b>Tabla 15:</b> Prueba LSD de Fisher por interacciones.....	59

## LISTA DE FIGURAS

<b>Figura 1:</b> Diseño de medidas repetidas	11
<b>Figura 2:</b> Tendencia de ganancia peso por semana de los tratamientos 1 y 2	46
<b>Figura 3:</b> Tendencia de ganancia de peso por semana de los 17 animales	46
<b>Figura 4:</b> Figura de normalidad Q-Q Plot	47
<b>Figura 5:</b> Residuos vs predichos, para la Homoscedasticidad de varianzas	50
<b>Figura 6:</b> Residuos vs semana, para la independencia	51
<b>Figura 7:</b> Prueba de normalidad	53
<b>Figura 8:</b> Residuos vs predichos, para la Homoscedasticidad de varianzas	53
<b>Figura 9:</b> Prueba de independencia	55
<b>Figura 10:</b> Figura LSD Fisher	60
<b>Figura 11:</b> Variabilidad por animal	61
<b>Figura 12:</b> Figura de normalidad	61
<b>Figura 13:</b> Figura de Homoscedasticidad	62

## RESUMEN

La presente tesis describe, sustenta y analiza las Medidas Repetidas aplicando análisis univariado, análisis multivariado y modelos lineales mixtos con el objetivo de ver las ventajas y desventajas de dichos enfoques. La metodología de la investigación utilizada para el desarrollo puede considerarse como exploratoria y descriptiva por cuanto en todo el proceso se indaga los individuos que hacen parte del estudio y sus mediciones a través del tiempo y/o espacio.

Para los diferentes análisis se usó un experimento realizado a 17 reses jóvenes cuyo fin era medir el crecimiento de las estas al proporcionarles dos tipos de tratamientos distintos por 12 semanas tomados del libro de *Biostatistics for animal science*.(2004). Finalmente se mostró que el mejor enfoque resultó ser el de los modelos mixtos.

**Palabras claves:** modelo Lineal general, Modelo Lineal mixto, medidas repetidas

## ABSTRACT

The thesis describes, supports and analyzes Repeated Measures through Mixed Linear Models. This research can be considered as exploratory and descriptive because we investigate in the measurements through time and / or space. the individuals who are part of our study, under specific conditions to be able to apply the mixed linear model; An experiment was carried out on 17 young cattle, the purpose was to measure the cattle's growth by providing two different treatments for 12 weeks taken from the book *Biostatistics for animal science*.

**Keywords:** General Linear Model, Mixed Linear Model, repeated measures.

## INTRODUCCIÓN

Uno de los diseños experimentales más utilizados en áreas como la agricultura, la medicina, entre otras, es el conocido diseño de medidas repetidas, que consiste en medir la variable respuesta de interés de una misma unidad experimental en diferentes puntos del tiempo o espacio. En este análisis se realizan varias mediciones a través del tiempo o espacio al individuo que hace parte del estudio bajo ciertas condiciones. Cuando nos enfrentamos a este tipo de diseño experimental, nos encontramos que los datos poseen correlación entre sí, por ello se debe utilizar el método estadístico adecuado, para que así se pueda obtener una mayor eficacia en el análisis de datos.

Para el ajuste de este tipo de diseño experimental se han usado varios métodos de análisis, entre ellos el análisis de varianza (ANOVA), el análisis multivariado (MANOVA), los modelos mixtos entre otros. Sin embargo, los métodos del ANOVA y MANOVA requieren el cumplimiento supuestos, entre ellos el de la independencia entre las observaciones; por tal razón, algunas veces se recurren a los modelos mixtos, ya que este método ha resultado muy útil en el estudio y análisis de datos correlacionados.

El tema central de nuestra investigación es encontrar el método que mejor se adapte a este tipo de datos, ya que en algunas ocasiones se cometen errores en el análisis, ocasionado por la mala utilización de ciertos métodos. Por ello se mostrará la teoría de los métodos de análisis univariado, análisis multivariado y modelos lineales mixtos. Adicionalmente se analizará un conjunto de datos bajo las tres metodologías para así mostrar las ventajas y desventajas en cada uno de los enfoques.

## 1. ANTECEDENTES

A continuación, se presentaran algunas investigaciones que facilitan la contextualización y desarrollo de las medidas repetidas, un ejemplo de ello es una investigación de Ruiz, M. (2004), cuyo objetivo principal consiste en dar una visión comparativa de los distintos métodos estadísticos como: (ANOVA, MANOVA Y MODELOS MIXTOS) desde un punto de vista práctico en el entorno de aplicaciones médicas, para cada uno de los métodos se plantea una breve introducción y una discusión de las condiciones de aplicación que permiten tener una idea del tipo de diseños en los que se puede utilizar y por último, se realizará el análisis de datos del resultado de un ensayo clínico donde se determina el estado de cada individuo a lo largo de tres instantes de tiempo ( $t_1$ ,  $t_2$  y  $t_3$ ), bajo dos factores (control y tratamiento), concluyendo que el inconveniente principal reside en la restricción de la matriz de covarianzas, lo que limita al método univariado, pero no tanto al método multivariado, requiriendo que los datos sean equilibrados y completos lo cual es muy difícil, por lo tanto, el último método utilizado fue el modelo mixto, el cual se ajusta bien a la matriz de correlación y a datos desequilibrados dando como conclusión que los modelos mixtos se ajustan correctamente a datos con medidas repetidas.

Las medidas repetidas son un diseño de modelo experimental interesante ya que con ellas se pueden usar gran variedad de métodos estadísticos como los expuestos por Gómez *et al.* (2012), cuyo objetivo principal es reseñar los procedimientos estadísticos que se usan en el diseño y análisis de medidas repetidas a través del tiempo, tales como: el análisis de varianza univariado y multivariado. Exponen que en estos dos métodos es muy común que los supuestos no se cumplan. Adicionalmente, indican cuales son los modelos de efectos fijos y los modelos de efectos aleatorios; nos explican cómo se constituyen los modelos mixtos, además, de cómo estos nos permiten analizar de forma correcta y eficientemente los datos de experimentos con medidas repetidas mediante el modelaje de la estructura de covarianzas, que consideran las correlaciones entre medidas repetidas y la presencia de varianzas heterogéneas. Por último, nos manifiesta cuáles son los criterios de selección de modelos mixtos, basados en los criterios de AIC



(Criterio de Información de Akaike) y BIC (Criterio de Información Bayesiano de Schwartz), indicando en qué contexto de la investigación es viable usar cada criterio.

Por otro lado, está la investigación de Mena, M. (2004), quien hace una revisión de las pruebas estadísticas alternativas que se usan cuando se viola el supuesto de esfericidad, incluyendo diferentes procedimientos de ajuste del estadístico  $f$ , el análisis multivariado, las pruebas de la aproximación general y de la aproximación general mejorada, el procedimiento de Welch-James, el acercamiento Bayesiano y el enfoque del modelo mixto.

Además, realiza una cuestión de estos procedimientos en términos de robustez estadística ante la violación de los supuestos para analizar los efectos principales y de interacción en diseños de medidas repetidas balanceadas y no balanceadas, concluyendo que con diseños balanceados, el uso de las pruebas  $F$  ajustadas son alternativas factibles ante la falta de esfericidad para comprobar los efectos principales y de interacción, el análisis multivariado solo resulta conveniente cuando hay un elevado número de sujetos, ya que no es robusto para la interacción si el número de sujetos es pequeño en relación con el número de medidas repetidas, pero en datos desbalanceados no es muy recomendado el uso generalizado de estas pruebas, ante los casos en los que no se recomienda las pruebas  $F$  ajustadas, es posible que los procedimientos de la Aproximación general mejorada y el de Welch-James se conviertan en alternativas adecuadas de análisis, no obstante el segundo requiere un mayor número de sujetos, en cuanto al enfoque del modelo mixto, se tiene el problema de la identificación correcta y adecuada de la estructura de las matrices de covarianzas.

Adicionalmente la investigación realizada por Fernández *et al.* (2007), se muestra que la potencia y robustez de los métodos estadísticos para el análisis de medidas repetidas están en función de la satisfacción de los supuestos asociados al análisis, en especial, el supuesto de esfericidad y de homogeneidad de las matrices de covarianza, expresando las técnicas univariadas (prueba de Greenhouse y Geisser, prueba de Huynh y Feldt, el enfoque de la aproximación general mejorada) y multivariadas (prueba

multivariada de Welch-James, prueba multivariada de Welch-James con estimadores robustos, prueba multivariada de Brown-Forsythe).

Como hemos visto, las medidas repetidas son un campo interesante para la utilización de diversos métodos y pruebas estadísticas, pero tan solo son una rama de los datos longitudinales; esta clase de datos es muy compleja a la hora de escoger el procedimiento adecuado. Gras, A. (2007), plantea por su parte una exploración de los principales procedimientos de análisis de datos longitudinales, haciendo hincapié en la clara distinción entre los procedimientos clásicos, basados en el análisis de las varianzas, y los procedimientos más actuales basados en los modelos de regresión, como el modelo general mixto.

## **2. JUSTIFICACIÓN**

En la actualidad esta es una de las alternativas de análisis de los datos longitudinales más usada, específicamente cuando es necesario ajustar los datos a las diferentes estructuras de matrices de variancia/covariancia; el contexto de esta investigación se basaba en los datos longitudinales con medidas repetidas realizando un énfasis en cada método estadístico, para mostrar cuál se ajusta más ante este diseño, así se llega a la conclusión de que los modelos estadísticos mixtos asumen que las observaciones constan de dos partes, los efectos fijos y los efectos aleatorios, los efectos fijos expresan los valores esperados de las observaciones, mientras que los efectos aleatorios reflejan las variancias y covariancias de las observaciones, lo cual es gran ventaja para modelar dichos diseños experimentales.

Esta investigación se realizó con el fin de comparar los diferentes métodos estadísticos utilizados en el análisis de medidas repetidas y mostrar sus debilidades y fortalezas a la hora de utilizarlos, para que así cuando nos enfrentemos a este tipo de datos usemos la herramienta estadística adecuada y no se comentan errores en las conclusiones del análisis de este tipo de datos.

### **3. OBJETIVOS**

#### **3.1 OBJETIVO GENERAL**

Mostrar los procedimientos y métodos de ajuste que se utiliza en un diseño de medidas repetidas a través del tiempo para dar a conocer sus deficiencias y fortalezas.

#### **3.2 OBJETIVOS ESPECÍFICOS**

- Contextualizar la teoría del análisis de medidas repetidas en el tiempo frente al ANOVA, MANOVA y los modelos mixtos.
- Analizar un conjunto de datos empíricos que muestre el ajuste de este tipo de diseño usando el software estadístico R-Project.
- Mostrar las ventajas y desventajas de los métodos usados para el estudio de datos con un diseño de medidas repetidas.
- Recomendar un método de análisis para el estudio de datos bajo un diseño de medidas repetidas en el tiempo.

#### 4. METODOLOGÍA

Los datos que se trabajaron en la investigación son extraídos de “*Biostatistics for Animal Science*<sup>1</sup>” el cual trae un estudio incorporado donde se midieron a 17 reses jóvenes con el fin de medir el crecimiento de las reses al proporcionarles dos tipos de tratamientos distintos por 12 semanas, dichos datos fueron recopilados a partir de la semana 9, 10, 11 y 12, generando así, un estudio de medidas repetidas a través del tiempo.

Los métodos estadísticos que van hacer parte del análisis son el análisis de varianza, el análisis multivariado y los modelos lineales mixtos. Se probarán y mostrarán las diferentes condiciones que imparten cada método, concluyendo de forma adecuada al respectivo resultado. Además, se darán a conocer las diferentes debilidades y fortalezas de cada método y se escogerá el mejor método que se adecue al respectivo análisis y ajustes de los datos.

Todos los análisis hechos se realizaron con el software de lenguaje de programación R-Project.

---

<sup>1</sup> Libro escrito en 2009 por Kaps, M., & Lamberson, W.

## **5. MARCO TEÓRICO**

En este capítulo se presenta la definición de dos categorías fundamentales para el desarrollo del trabajo, una es medidas repetidas y la otra modelos lineales mixtos. Además, se hace una contextualización de las metodologías usadas en el análisis de diseños con medidas repetidas tales como: Análisis univariado (ANOVA), análisis multivariado (MANOVA) y modelo lineal mixto (MLM).

A continuación, los planteamientos teóricos que se presentan (Littell *et al.* 2006).

### **5.1. INTRODUCCIÓN A LAS MEDIDAS REPETIDAS**

El término medidas repetidas se refiere a conjuntos de datos con múltiples mediciones de una variable de respuesta en la misma unidad experimental. En la mayoría de las aplicaciones, las múltiples mediciones se realizan durante un período de tiempo. Un ejemplo son los datos de la curva de crecimiento, como las mediciones mensuales de peso de los bebés durante el primer año de vida. Otro ejemplo son los datos sobre los efectos de los fármacos, tales como las mediciones del pulso o la respiración en los pacientes después de la administración de un fármaco. Pero las medidas repetidas también pueden referirse a múltiples mediciones sobre el espacio, tales como espesores de las vértebras de los animales. En un sentido general, los datos que se miden repetidamente en el tiempo o en el espacio son datos de medidas repetidas.

La mayor parte de este capítulo usa el término en el sentido más tradicional, haciendo referencia a secuencias de medidas en unidades experimentales en un experimento diseñado, encuesta por muestreo o estudio retrospectivo.

El diseño de medidas repetidas utiliza los mismos sujetos con todas las condiciones de la investigación, para aplicarles tratamientos mediando el control; requiriendo con esto un número menor de participantes y recursos en donde también se debe disminuir los

efectos de la variación natural entre los individuos sobre los resultados. Los diseños de sujetos repetidos se utilizan comúnmente en los estudios de medidas repetidas, con un largo plazo, y en las pruebas educativas en donde es importante asegurar que la variabilidad sea baja.

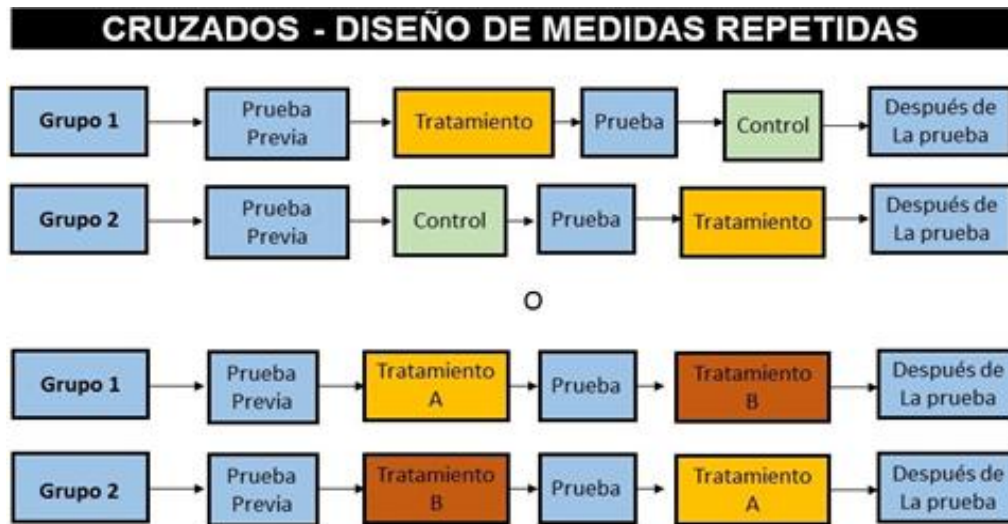
El diseño de medidas repetidas es actualmente un proceso para la recolección de información utilizado principalmente en el campo de las ciencias de la salud, las ciencias sociales y también en la psicología. Además, se puede decir que:

Cuando administramos los tratamientos objeto de nuestra investigación a los mismos sujetos y, en consecuencia, estos reciben más de un tratamiento experimental, disponiendo al menos de una observación por tratamiento y sujeto, decimos que estamos en presencia de un diseño intra-sujetos o de medidas repetidas. (Pascual.*et al.*1996)

Los diseños de medidas repetidas se pueden utilizar en diferentes tipos de situaciones (Maxwell y Delaney, 1990).

- a)** Evaluación longitudinal del cambio a lo largo del tiempo.
- b)** Evaluación de la actuación de los sujetos bajo diferentes condiciones de tratamiento en estudios transversales.
- c)** En aquellas situaciones en las que se desea comparar las puntuaciones de los mismos sujetos obtenidas en diferentes pruebas psicométricas.

**Figura 1:** *Diseño de medidas repetitivas*



Fuente: recuperado de: <https://explorable.com/es/disenio-de-medidas-repetidas>

Al igual que todos los diseños de medidas repetidas, se reduce a la posibilidad de variación entre los individuos desviando los resultados y también requiere un grupo más pequeño de sujetos. Asimismo, reduce la posibilidad de efectos de la práctica que influyen en los resultados, ya que, presumiblemente, será el mismo para ambos grupos y puede ser eliminado mediante pruebas estadísticas. El mayor escollo es si los efectos de arrastre son asimétricos, por ejemplo, si  $B$  afecta a  $A$  más que  $A$  a  $B$  (ver grafica 1).

**5.1.1** Conceptos básicos de medidas repetidas. Un estudio de medidas repetidas de uso frecuente consiste en un diseño experimental completamente al azar con datos recogidos en una secuencia de puntos de tiempo equidistantes de cada unidad experimental. Gran parte del desarrollo de la metodología de medidas repetidas se produjo en el área de la psicología humana. Como resultado, las unidades experimentales a menudo se denominan sujetos. Pero "sujeto" podría referirse a un animal, una muestra de laboratorio, o una pieza de equipo industrial.



En esta disposición básica de un diseño completamente al azar con medidas repetidas, hay dos factores, tratamientos y tiempo. En este sentido, todos los experimentos de medidas repetidas son experimentos factoriales. El tratamiento se denomina factor entre sujetos porque los niveles de tratamiento pueden cambiar sólo entre sujetos; Todas las mediciones sobre el mismo sujeto representarán el mismo tratamiento. El tiempo se denomina un factor dentro de los sujetos porque se toman diferentes medidas sobre el mismo sujeto en momentos diferentes. En los experimentos de medidas repetidas, el interés se centra en **(1)** cómo los medios de tratamiento difieren, **(2)** cómo los medios de tratamiento cambian con el tiempo, y **(3)** cómo las diferencias entre los medios de tratamiento cambian con el tiempo. En otras palabras, ¿hay un efecto principal del tratamiento, hay un efecto principal del tiempo, y hay una interacción del tratamiento-por-tiempo? Estos son los tipos de preguntas que queremos hacer en cualquier estudio de dos factores. Principalmente, la interacción sería la primera pregunta a investigar.

No hay nada peculiar en los objetivos de un estudio de medidas repetidas. Lo que hace que el análisis de datos de medidas repetidas sea distinto es la estructura de covarianza de los datos observados. En los diseños de bloques al azar, los tratamientos se asignan al azar a unidades dentro de un bloque. Esto hace que todas las observaciones dentro de un bloque dado estén igualmente correlacionadas. Sin embargo, en experimentos de medidas repetidas, dos medidas tomadas en puntos de tiempo adyacentes son típicamente más altamente correlacionadas que dos mediciones tomadas varios puntos de tiempo separados. El esfuerzo es usualmente necesario al comienzo del análisis estadístico para evaluar la estructura de covarianza de los datos. El modelado de una estructura de covarianza apropiada es esencial para que pueda hacerse una inferencia válida en la forma de pruebas de hipótesis e intervalos de confianza sobre los medios de tratamiento.

Existen similitudes entre los experimentos de medidas repetidas y los experimentos con parcelas subdivididas. El factor de tratamiento en un experimento de medidas repetidas corresponde al factor de principal en un experimento de parcelas subdivididas. El factor tiempo en medidas repetidas corresponde al factor de sub. En otras palabras, el factor

entre sujetos corresponde al factor de trazado principal, y el factor dentro de los sujetos corresponde al factor de sub.

Las unidades experimentales a las que se asignan los tratamientos en el experimento de medidas repetidas son análogas a las unidades principales en el experimento de parcelas subdivididas y las unidades experimentales en momentos particulares corresponden a unidades de sub. Sin embargo, en un verdadero experimento de parcelas divididas, los niveles del factor de sub se asignan aleatoriamente a unidades de subparcelas dentro de las unidades de la parcela principal. En consecuencia, las respuestas

De diferentes unidades de subparcelas en la misma unidad de parcela principal están correlacionadas de manera similar entre sí. Pero en experimentos de medidas repetidas, las respuestas de puntos cercanos en el tiempo suelen estar más altamente correlacionadas que las respuestas de puntos distanciados en el tiempo. Por lo tanto, generalmente se necesitan métodos especiales de análisis para acomodar la estructura de correlación de las medidas repetidas.

Consideremos una situación experimental en la que los sujetos son asignados aleatoriamente a un factor de tratamiento, y las mediciones se realizan a intervalos equitativos en cada sujeto. Sea  $Y_{ijk}$  la medida en el tiempo  $k$  sobre el  $j$  – ésimo sujeto asignado al tratamiento  $i$ .

### Modelo

Un modelo estadístico para datos de medidas repetidas es:

$$Y_{ijk} = \mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} + e_{ijk} \quad (1)$$

### Donde

$\mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik}$ : Es la media del tratamiento  $i$  en el tiempo  $k$ , que contiene efectos para el tratamiento, el tiempo y el tratamiento x tiempo de interacción

$e_{ijk}$ : es el error aleatorio asociado con la medición en el momento  $k$  sobre el  $j$  – ésimo sujeto Que se asigna al tratamiento  $i$

La ecuación del modelo anterior es la misma que la ecuación del modelo para un experimento factorial estándar con los principales efectos del tratamiento y del tiempo y la interacción del tratamiento  $\times$  tiempo. La característica distintiva de un modelo de medidas repetidas es la estructura de varianza y covarianza de los errores  $e_{ijk}$ . Aunque los tratamientos se asignaron aleatoriamente a sujetos, los niveles del factor de medidas repetidas, en este caso el tiempo, no se asignan aleatoriamente a unidades dentro de los sujetos. Por lo tanto, no podemos suponer razonablemente que los errores aleatorios  $e_{ijk}$  para el mismo sujeto son independientes. En su lugar, suponemos que los errores para diferentes sujetos son independientes, dando:

$$Cov[e_{ijk}, e_{i'j'l'}] = 0 \text{ ya sea } i \neq i' \text{ o } j \neq j' \quad (2)$$

Además, dado que la medición sobre el mismo sujeto es sobre un lapso de tiempo, pueden tener diferentes varianzas, y las correlaciones entre los pares, de las mediciones puede depender de la longitud del intervalo de tiempo entre las mediciones. Por lo tanto, en el ajuste más general, sólo asumimos

$$Var[e_{ijk}] = \sigma_k^2 \text{ y } Cov[e_{ijk}, e_{ijk}] = \sigma_{kk} \quad (3)$$

En otras palabras, permitimos que la varianza de  $e_{ijk}$  dependa del tiempo de medición  $k$ , y la covarianza entre los errores a dos tiempos,  $k$  y  $k'$ , para el mismo sujeto, depende de los tiempos. En la mayoría de los casos, el modelo para la covarianza puede expresarse de acuerdo con una estructura que implica menos parámetros. Si se expresa el vector de observaciones sobre el sujeto  $j$  en el tratamiento  $i$  como  $Y_{ij} = [y_{ij1}, \dots, y_{ijl}]'$ , entonces tenemos  $Var[Y_{ij}] = \Sigma$ , donde el elemento en la fila  $k$  y la columna  $k'$  es  $\sigma_{kk'}$ . Esto supone que la matriz de covarianza  $\Sigma$  es la misma para todos los sujetos. Si se apilan los vectores  $Y_{ij}$  en un solo vector  $Y = [Y'_{11}, Y'_{12}, \dots, Y'_{tn}]'$ , entonces  $Var[Y] = V$  es el bloque diagonal con  $\Sigma$  a lo largo de la diagonal. Podemos escribir  $\Sigma$  como

$$\text{Var} [Y] = \mathbf{I}_k \otimes \Sigma(4)$$

Donde  $\mathbf{I}_k$  es una matriz de identidad de dimensión igual al número de sujetos. En algunas situaciones, es ventajoso incluir un efecto aleatorio entre sujetos para dar el modelo

$$Y_{ijk} = \mu + \alpha_i + b_{ij} + \gamma_k + (\alpha\gamma)_{ik} + \varepsilon_{ijk}(5)$$

Donde  $b_{ij}$  es un efecto aleatorio para el sujeto  $j$  asignado al tratamiento  $i$ , y  $\varepsilon_{ijk}$  es un error con matriz de covarianza  $R$  con una estructura paramétrica. La matriz de covarianza de  $Y_{ij} = [Y_{ijl}, \dots, Y_{ijl}]'$  se convierte en

$$\Sigma = \text{Var}[Y_{ij}] = \sigma_b^2 J + R(6)$$

Donde  $J$  es una matriz de unos. La ecuación muestra los dos aspectos de la covarianza entre medidas sobre el mismo sujeto. La parte  $\sigma_b^2 J$  representa la covarianza debido al hecho de que las medidas están sobre el mismo sujeto, y  $R$  representa la contribución a la covarianza debido a la proximidad de las mediciones.

El modelo es similar a los modelos de tipo parcela dividida, donde  $b_{ij}$  corresponde a errores de parcela completa y  $\varepsilon_{ijk}$  corresponde a errores de sub-parcela. La distinción es que el  $\varepsilon_{ijk}$ , que correspondería a los errores de la subtrama, no puede ser necesariamente asumido independiente e idénticamente distribuido porque los efectos del tiempo pueden no estar correlacionados igualmente dentro de los sujetos.

Los datos de medidas repetidas se pueden analizar utilizando métodos de modelo mixto basados en mínimos cuadrados generalizados y máxima verosimilitud. En la notación matricial del modelo es,

$$Y = X\beta + e \quad (7)$$

Donde  $Y$  es el vector de datos observados,  $X$  es una matriz de constantes conocidas,  $\beta$  es un vector de parámetros fijos pero desconocidos, y  $e$  es un vector de errores aleatorios con matriz de covarianza  $\text{Var} [e] = V$ .

Entonces la expectativa del vector de observaciones es  $E[Y] = X\beta$  y su varianza es  $Var[Y] = V$ . El estimador  $\beta$  es

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad (8)$$

La matriz de covarianza de  $\hat{\beta}$  Es

$$Var[\hat{\beta}] = (X'V^{-1}X)^{-1} \quad (9)$$

En la situación de esta sección,  $V^{-1} = I_k \otimes \Sigma$ , que muestra más directamente por qué es necesario modelar la forma de  $\Sigma$ , la matriz de covarianza para un sujeto individual.

Los datos que permiten analizar este modelo son los procedentes de un diseño con un solo grupo de sujetos y un único factor cuyos niveles se aplican a todos los sujetos. Las distintas medidas, tantas como niveles tiene el factor, se toman sobre los mismos sujetos; de ahí el nombre de medidas repetidas que reciben estos modelos.

Ahora, los planteamientos teóricos que se presentan son los de (Falcó, M .2009).

## 5.2. ANÁLISIS DE VARIANZA UNIVARIADO (ANOVA)

La técnica del Análisis de la Varianza (ANOVA) es una de las técnicas más utilizadas en el análisis de datos de los diseños experimentales. Se utiliza cuando se quiere probar la igualdad de dos o más medias, por lo que puede verse como una extensión de la prueba  $t$  para diferencias de dos medias. Un análisis de la varianza permite determinar si diferentes tratamientos muestran diferencias significativas o por el contrario puede suponerse que sus medias poblacionales no difieren. El ANOVA también puede verse como un procedimiento que permite dividir la varianza de la variable dependiente en dos o más componentes, cada una de las cuales puede ser atribuido a una fuente (variable o factor) identificable.

El modelo a una vía de clasificación utilizado por el ANOVA es el siguiente:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \text{ variando } i = 1, 2, \dots, n \text{ y } j = 1, 2, \dots, t \quad (10)$$

Dónde:

$Y_{ij}$  : Variable dependiente.

$\mu$  : Media general del modelo.

$\tau_i$  : El efecto del  $i$  – *esimo* tratamiento.

$\varepsilon_{ij}$ : El error aleatorio independiente e idénticamente distribuido como una normal con media 0 y varianza  $\sigma^2$ ,  $\varepsilon_{ij} \sim N(0, \sigma^2)$ . (11)

### 5.2.1. Tabla ANOVA

Denominando  $S$  a la suma de los cuadrados, se tiene:

$$S_R^2 = \frac{S_R}{N-K} \quad (12)$$

$$S_T^2 = \frac{S_T}{K-1} \quad (13)$$

Si  $S_D$  es la suma de los cuadrados con respecto a la media global, el estadístico  $S_D^2$  es también un estimado de  $\sigma^2$  si se cumplen las hipótesis de igualdad de medias:

$$S_D^2 = \frac{\sum_i^K \sum_j^{n_i} (y_{ij} - \bar{y})^2}{N-1} = \frac{S_D}{N-1} \quad (14)$$

Es fácil comprobar que se verifica la siguiente igualdad:

$$S_D = S_R + S_T \quad (15)$$

En cuanto a los grados de libertad:

$$V_D = V_R + V_T, \text{ es decir } N - 1 = (N - K) + (K - 1) \quad (16)$$

$S_D$  Se denomina también “suma corregida de cuadrados” y se calcula fácilmente mediante la siguiente ecuación:

$$S_D = \sum_i^K \sum_j^{n_i} y_{ij}^2 - N\bar{y}^2 \quad (17)$$

En general lo más cómodo es calcular  $S_D$  y  $S_T$ , calculando  $S_R$  por diferencia. Es costumbre presentar el ANÁLISIS DE LA VARIANZA en forma de tabla:

**Tabla 1.** Análisis de la Varianza 1 factor

FUENTE DE VARIACIÓN	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	CUADRO MEDIO	CONTRA STE
ENTRE TRATAMIENTOS (VE)	$S_T$ $= \sum_i^k n_i(\bar{Y}_{ti} - \bar{Y})$	$V_T = K - 1$	$S_T^2$ $= \frac{S_T}{K - 1}$	$\frac{S_T^2}{S_R^2}$
DENTRO DE TRATAMIENTOS (VNE)	$S_R$ $= \sum_i^k \sum_j^{n_i} (y_{ij} - \bar{y}_i)^2$	$V_R = N - K$	$S_R^2$ $= \frac{S_R}{N - K}$	$\frac{S_T^2}{S_R^2}$
TOTAL, EN RELACIÓN A LA MEDIA GENERAL (VT)	$S_D$ $= \sum_i^k \sum_j^{n_i} y_{ij}^2 - N\bar{y}^2$	$V_D = N - 1$	$S_D^2$ $= \frac{S_D}{N - 1}$	

### 5.2.2. Supuestos.

Para que se pueda aplicar el ANÁLISIS DE LA VARIANZA es preciso que se cumplan estas tres hipótesis:

- Los datos han de ser independientes. Para asegurar esto, las muestras cuyas medias se desea comparar han de extraerse de manera aleatoria.
- Las poblaciones base de donde proceden las muestras han de ser normales.
- Las poblaciones base de donde proceden las muestras han de tener la misma varianza (Heteroscedasticidad). Estas hipótesis implican que las perturbaciones se distribuyan según una  $N(0, \sigma^2)$ .

Los supuestos en los que está basado el ANOVA se resumen en los siguientes:

1. Los valores esperados de los errores son ceros.
2. Las varianzas de todos los errores son iguales entre sí.

3. Los errores son independientes.
4. Los errores se distribuyen normalmente, con media cero y varianza  $\sigma^2$ .

Los supuestos anteriormente mencionados, se pueden verificar usando diversas técnicas estadísticas, por ejemplo para probar la normalidad se pueden usar las técnicas Figuras como el *qq – norm* y algunos test como el de *Kolmogorov – Smirnov* y *Shapiro – Wilk*; mientras que para la Homoscedasticidad de varianzas se usa el test de *Levene* o de *Bartlett*.

### 5.3. ANÁLISIS DE VARIANZA MULTIVARIADO (MANOVA)

Los planteamientos teóricos que se presentan en (Díaz, L.2002).

La distinción entre los modelos lineales multivariados y los modelos univariados, es que en el multivariado se involucra más de una variable dependiente o respuesta. Considérese que las observaciones multivariadas  $Y_1, \dots, Y_n$ , conforman un conjunto de observaciones independientes de una población normal  $p$ -variante, es decir,  $Y_\alpha \sim N_p(X_\alpha \beta, \Sigma)$ , para  $\alpha = 1, \dots, n$ . En este caso utilizamos un vector  $X_\alpha$  de tamaño  $(1 \times q)$  el cual conocemos, pero la matriz  $\sigma_{p \times p}$  como la matriz  $\beta_{q \times p}$  son desconocidas.

Los  $Y_\alpha$  corresponden a las variables respuesta en un modelo de regresión (dependientes), mientras que las  $X_\alpha$  son las variables regresoras o explicativas. En tales condiciones los vectores se pueden relacionar a través de un modelo lineal general multivariado, tal como el siguiente:

$$\begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_{11} & \cdots & \beta_{1p} \\ \vdots & \ddots & \vdots \\ \beta_{n1} & \cdots & \beta_{np} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \cdots & \epsilon_{1p} \\ \vdots & \ddots & \vdots \\ \epsilon_{n1} & \cdots & \epsilon_{np} \end{pmatrix} \quad (18)$$

En forma más condensada, el modelo lineal multivariado anterior se escribe de la manera siguiente:



$$Y = X\beta + \varepsilon, \quad (19)$$

Donde  $X$  conforma, en la mayoría de los casos, la matriz de diseño o la matriz de variable regresoras,  $\beta$  es la matriz de parámetros desconocidos y la matriz aleatoria  $\varepsilon$  contiene los errores. Los estimadores de máxima verosimilitud para  $\beta$  y  $\sigma$  son:

$$\hat{\beta} = \left( \sum_{\alpha=1}^n X'_{\alpha} X_{\alpha} \right)^{-1} \left( \sum_{\alpha=1}^n X'_{\alpha} Y_{\alpha} \right)$$

$$\hat{\sigma} = \frac{1}{n} \sum_{\alpha=1}^n (Y_{\alpha} - X_{\alpha} \hat{\beta})(Y_{\alpha} - X_{\alpha} \hat{\beta})' \quad (20)$$

Observaciones:

- a. Con estos estimadores podemos deducir lo correspondiente a la regresión lineal múltiple, donde  $q = 1$ . El estimador de máxima verosimilitud  $\hat{\beta}$ , dado en (2) tiene distribución normal, con vector de medias  $\beta$ , matriz de varianzas y covarianzas que resultan del producto directo entre  $\sigma$  y  $A^{-1}$ ; es decir,

$$Cov(\hat{\beta}) = \sigma \otimes A^{-1} = \begin{pmatrix} \sigma_{11}A^{-1} & \sigma_{12}A^{-1} & \dots & \sigma_{1p}A^{-1} \\ \sigma_{21}A^{-1} & \sigma_{22}A^{-1} & \dots & \sigma_{2p}A^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1}A^{-1} & \sigma_{p2}A^{-1} & \dots & \sigma_{pp}A^{-1} \end{pmatrix} \quad (21)$$

Donde,

$$A = \sum_{\alpha=1}^n X'_{\alpha} X_{\alpha} \quad (22)$$

- b. Se nota la similitud con el modelo de regresión lineal, donde se asume que los errores tienen matriz de covarianzas  $\sigma = \sigma^2 I$ , así que  $Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1}$  es un caso especial de la última expresión.
- c. De manera similar, el estimador de máxima verosimilitud  $n\hat{\sigma}$  es distribuido normalmente como  $W(\sigma, n - q)$ , e independiente de  $\hat{\beta}$ , con  $q$  el número de componentes de  $X_{\alpha}$ .
- d. Para obtener un estimador insesgado de  $\sigma$  se debe hacer  $S = \left( \frac{n}{(n-q)} \right) \hat{\sigma}$ .

### 5.3.1. Contrastes de hipótesis

Suponemos que se particiona la matriz de parámetros  $\beta$  como:

$$\beta = (\beta_1 : \beta_2) \quad (23)$$

Como  $\beta_1$  de la columna  $q_1$  y  $\beta_2$  de la columna  $q_2$  se tiene que  $(q_1 + q_2 = q)$ . Y con esto probar hipótesis con la razón de máxima verosimilitud.

$$H_0 : \beta_1 = \beta_1^* \quad (24)$$

Se obtiene en forma semejante a como se realiza con la estadística  $T^2$ ; esta es,

$$\lambda = \frac{|\hat{\beta}_1|^{n/2}}{|\hat{\beta}_2|^{n/2}} \quad (25)$$

La matriz  $\hat{\beta}$  corresponde al estimador máximo verosímil en el espacio global de los parámetros.

**5.3.2. Modelos de una vía de clasificación.** Desde el punto de vista práctico, el MANOVA es una técnica con la cual se puede verificar la igualdad de los vectores de medias ligados a varias poblaciones multivariadas (Díaz, 2007). De esta manera, se presenta a continuación la técnica de análisis de varianza para arreglos de una vía de clasificación. Consideremos que  $Y_{ij}$  es una observación de una población  $N_p(\mu_i, \sigma)$  con  $i = 1, \dots, n_i$  y  $j = 1, \dots, q$ . Los datos se pueden visualizar de la siguiente forma:

<b>Población</b>	<b>Muestra</b>	<b>Media Muestral</b>
<i>Pob. 1</i>	$Y_{11}, Y_{12}, \dots, Y_{1n_1}$	$\hat{Y}_1$
<i>Pob. 2</i>	$Y_{21}, Y_{22}, \dots, Y_{2n_2}$	$\hat{Y}_2$
$\vdots$	$\vdots$	$\vdots$
<i>Pob. q</i>	$Y_{q1}, Y_{q2}, \dots, Y_{qn_q}$	$\hat{Y}_q$

(26)

Nótese que se han considerado  $n_i$  observaciones en cada población, este es el caso más general. Si los  $n_i$  son diferentes se dice que se trata de un diseño experimental desbalanceado; cuando  $n_i = \dots = n_q = n$  se dice que ello diseño es balanceado.

La media  $\hat{Y}_i$  en cada muestra se obtiene mediante,

$$\hat{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_i} Y_{i.} \quad \text{Para } i = 1, \dots, q \quad (27)$$

La media general  $\hat{Y}_{..}$  se obtiene de:

$$\hat{Y}_{..} = \frac{1}{N} \sum_{i=1}^q \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{N} \sum_{i=1}^q \hat{Y}_i \quad (28)$$

Con  $N = \sum_{i=1}^q n_i$ , el número total de observaciones.

El modelo que relaciona las observaciones con los parámetros  $\mu_i$  es de la forma:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \text{ con } \varepsilon_{ij} \sim N_p(0, \sigma), \text{ para } i = 1, \dots, q \text{ y } j = 1, \dots, n_i \quad (29)$$

Las hipótesis a verificar es la igualdad de los vectores de medias de la q-población; es decir:

$$H_0 : \mu_1 = \dots = \mu_q \quad (30)$$

Una expresión equivalente con (4) es:

$$\beta_1 = (\mu_1 - \mu_q, \dots, \mu_{q-1} - \mu_q) \quad (31)$$

$$\beta_2 = \mu_q$$

La hipótesis planteada en (5) se puede escribir en la forma,

$$H_0 = \mu_1 - \mu_q = \mu_2 - \mu_q = \dots = \mu_{q-1} - \mu_q = 0 \quad (32)$$

$$: \begin{pmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix} \quad (33)$$

La región de rechazo a un nivel de significancia  $\alpha$  es:

$$\Lambda = \frac{|E|}{|E+H|} = \frac{|N\hat{\sigma}|}{|N\hat{\sigma}_\omega|} < \Lambda_{(\alpha, p, v_H, v_E)} \quad (34)$$

Donde  $v_H = q - 1$  son los grados de libertad para la hipótesis,  $v_E = N - q$  son los grados de libertad del error ( $N = \sum_{i=1}^q n_i$ ).

#### 5.4. MEDIDAS REPETIDAS EN Q-MUESTRAS

El diseño de medidas repetidas implica un modelo de una sola vía de clasificación:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \dots, n \text{ y } j = 1, \dots, t \quad (35)$$

Donde,

$Y_{ij}$ : La puntuación del  $i$ -ésimo sujeto bajo la  $j$ -ésima condición experimental o tratamiento

$\mu_i$  : La media asociado al efecto del  $i$ -ésimo sujeto

$\varepsilon_{ij}$ : El error experimental asociado al  $i$  - ésimosujeto bajo el  $j$  - ésimotratamiento

Desde los  $q$ -grupos de  $n$ -observaciones cada uno, se calcula  $\bar{Y}_1 \dots \bar{Y}_q$  y la matriz de errores  $\varepsilon$ . Los datos se ubican en una tabla de factores de  $A$  columnas y  $B$  filas, donde se consideran tres casos importantes para el análisis:

- El primero tiene en cuenta los niveles del factor  $B$  (filas) como un grupo o una población y se hace un análisis de las medidas repetidas ante los niveles del factor  $A$  (columnas).
- Se realiza un análisis entre los niveles del factor  $B$  (filas) para determinar su efecto en la variable respuesta. Los factores solo pueden asumir un número limitado de posibles valores, conocidos como niveles de factores.

- c. Luego se realiza un tercer análisis, para verificar las interacciones entre los efectos columnas  $A$  y las filas  $B$ .

Si la interacción resulta significativa, se debe proceder a la “apertura” de la misma. Esto consiste en “probar las hipótesis correspondientes a los efectos simples de uno de los factores para cada uno de los niveles del otro factor” (Cox, 1958). Considerando como término de error el correspondiente al modelo completo (Steel & Torrie, 1988). Posteriormente, para cada efecto simple que resulta significativo se debe aplicar una prueba de comparaciones múltiples (Prueba de Tukey, por ejemplo) para detectar grupos de niveles homogéneos, afectando sus niveles de probabilidad según el criterio de Bonferroni (Underwood, 1997). Con esto se puede lograr un análisis de aquellos que se logran en uno de doble vía de clasificación.

**Tabla 2.** Medidas repetidas en q- grupos

Factor B	Factor A (Medidas repetidas)				
	Sujeto		$A_1 A_2$	$\cdots$	$A_p$
	Grupos				
	$B_1 S_{11} (Y_{111} Y_{112}$	$\cdots$	$Y_{11p}) = Y'_{11}$		
	$S_{12} (Y_{121} Y_{122}$	$\cdots$	$Y_{12p}) = Y'_{12}$		
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$S_{1n} (Y_{1n1} Y_{1n2}$	$\cdots$	$Y_{1np}) = Y'_{1n}$		
	$B_2 S_{21} (Y_{211} Y_{212}$	$\cdots$	$Y_{21p}) = Y'_{21}$		
	$S_{22} (Y_{221} Y_{222}$	$\cdots$	$Y_{22p}) = Y'_{22}$		
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$S_{2n} (Y_{2n1} Y_{2n2}$	$\cdots$	$Y_{2np}) = Y'_{2n}$		

$$\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \vdots \\
& \vdots & & & & & \\
\hline
& B_q S_{q1} (Y_{q11} Y_{q12} & \cdots & Y_{q1p}) = Y'_{q1} \\
& S_{q2} (Y_{q21} Y_{q22} & \cdots & Y_{q2p}) = Y'_{q2} \\
& \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
& S_{qn} (Y_{qn1} Y_{qn2} & \cdots & Y_{qnp}) = Y'_{qn}
\end{array}$$

Para comprobar las medias de los  $q$  – *niveles* del factor  $B$ , se toman las medias en cada grupo. Éstas son el promedio sobre cada uno de los niveles del factor  $A$ ; es decir,  $\sum_{j=1}^p \frac{\mu_{ij}}{p} = \frac{1' \mu_i}{p}$ . La hipótesis se escribe como:

$$H_0: 1' \mu_1 = \cdots = 1' \mu_q \quad (36)$$

La cual es equivalente a probar que los perfiles fila están en el mismo nivel. Las expresiones  $1' \mu_i$ , para  $i = 1; \dots, q$  son escalares, luego esta hipótesis puede verificarse mediante la estadística  $F$ , como un análisis de varianza univariado a una vía de clasificación sobre  $Z_{ij} = 1Y_{ij}$ , para  $i = 1; \dots, q$  y  $j = 1; \dots, n_i$ . De esta manera, a cada sujeto  $S_{ij}$  se le hace corresponder el escalar  $Z_{ij}$ . Es decir, cada observación vectorial para cada sujeto o individuo se reduce a una observación de tipo escalar, luego, mediante un análisis de varianza univariado se comparan las medias  $1'; \dots; 1' \bar{Y}_q$ .

La hipótesis sobre la interacción  $AB$  es equivalente a la hipótesis de “paralelismo

$$H_0: C\mu_1 = \cdots = C\mu_q. \quad (37)$$

Así, las diferencias o contrastes entre los niveles del factor  $A$  son los mismo a través de los niveles del factor  $B$ , Este resultado se prueba fácilmente mediante un análisis de varianza multivariado a una vía de clasificación sobre  $Z_{ij} = CY_{ij}$ , con

$$\Lambda = \frac{|CEC'|}{|C(E+H)C'|}, (38)$$

La cual se distribuye  $\Lambda_{(p-1, q-1, N-q)}$ .

Observación:

El cálculo de las estadísticas de prueba para medidas repetidas puede hacerse mediante las matrices  $H$  y  $E$  del análisis de varianza multivariado. Otra forma consiste en transformar los datos de acuerdo con  $Z_{ij} = CY_{ij}$ . Para la hipótesis  $H_0$  asociada al factor  $A$ , por ejemplo, para  $p = 4$ ,

$$C = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} (39)$$

Así, cada observación de  $Y' = (Y_1, Y_2, Y_3, Y_4)$  se transforma por medio de  $Z' = (Y_1 - Y_2, Y_2 - Y_3, Y_3 - Y_4)$ . De esta forma se verifica la hipótesis  $H_0: \mu_Z = 0$  mediante la estadística para una muestra

$$T^2 = N\bar{Z}'S_Z^{-1}\bar{Z} (40)$$

Con  $N = \sum_{i=1}^q n_i$ ,  $\bar{Z} = \sum_{ij} \frac{Z_{ij}}{N}$  y  $S_Z = \frac{E_Z}{N-q}$ . Se rechaza la hipótesis  $H_0$  si  $T^2 \geq T_{(\alpha, p-1, N-q)}^2$ .

Para verificar la hipótesis  $H_0: C\mu_1 = \dots = C\mu_q$ , en el factor  $B$ , se suman las componentes de cada vector de observaciones, se obtiene

$$Z_{ij} = 1'Y_{ij} = Y_{ij} + \dots + Y_{ijp}, (41)$$

Luego se comprueban las medias  $\bar{Z}_1, \dots, \bar{Z}_q$  mediante una estadística  $F$  en un análisis de varianza univariado a una vía de clasificación.

Para la hipótesis  $Y_{ij} = \mu + \gamma_i + \varepsilon_{ij} = \mu_i + \varepsilon_{ij}$ , de interacción entre los factores  $A$  y  $B$ , se transforma cada  $Y_{ij}$  en  $Z_{ij} = CY_{ij}$ , empleando las filas de la matriz  $C$  anterior, El vector

$Z_{ij}$  resultante es un vector de tamaño  $(p - 1) * 1$ . Así, se debe hacer un análisis de varianza multivariado sobre  $Z_{ij}$  para obtener

$$\Lambda = \frac{|E_Z|}{|E_Z + H_Z|} \quad (42)$$

## 5.5. MODELO DE DOS FACTORES DENTRO DE UN SUJETO Y UN FACTOR ENTRE SUJETOS

Este modelo corresponde a una vía de clasificación multivariada que permite que cada vector de observaciones tenga medidas de un arreglo de tratamientos del tipo factorial de dos vías. Cada sujeto recibe su tratamiento, en este caso es la combinación de los niveles de los factores  $A$  y  $B$ .

En un diseño de dos factores, ambos con medidas repetidas, los sujetos que participan en el experimento pasan por todas las condiciones experimentales, es decir, por todas las condiciones definidas por las posibles combinaciones entre los niveles de ambos factores. Pulido *et al.* (2012)

Supongamos que tenemos 9 sujetos a los cuales se les aplicará un tratamiento, en este caso el arreglo quedaría así:

$$A_1 B_1 A_1 B_2 A_1 B_3 A_2 B_1 A_2 B_2 A_2 B_3 A_3 B_1 A_3 B_2 A_3 B_3 \quad (43)$$

En este caso se observa algo muy parecido a lo que se encuentra al probar hipótesis en diseños de “parcelas divididas” pero esta vez en forma multivariada. El modelo que esta observación toma es:

$$Y_{ij} = \mu + \gamma_i + \epsilon_{ij} = \mu_i + \epsilon_{ij} \text{variando } i = 1, 2, \dots, n \text{ y } j = 1, 2, \dots, t \quad (44)$$

**Donde,**



$Y_{ij}$ : Contrastes entre  $i$  – ésimo,  $j$  – ésimo tratamiento

$\mu$ : La media general

$\gamma_i$ : Es el efecto debido al  $i$  – ésimo nivel del factor  $C$

**Tabla 3.** Factores dentro de sujetos (A y B)

Factores dentro de sujetos (A y B)													
Entre	$\underbrace{\hspace{1.5cm}}^A$		$\underbrace{\hspace{1.5cm}}^A$		$\underbrace{\hspace{1.5cm}}^A$								
			Suj.										
(C)			Obs.	$B_1 B_2 B_3$	$B_1 B_2 B_3$		$B_1 B_2 B_3$						
<hr/>													
	$C_1 Y_{11} = (Y_{111} Y_{112} Y_{113} Y_{114} Y_{115} \ Y_{116} Y_{117} Y_{118} Y_{119})$												
	$Y_{12} = (Y_{121} Y_{122} Y_{123} Y_{124} Y_{125} \ Y_{126} Y_{127} Y_{128} Y_{129})$												
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$						
$\vdots$	$\vdots$												
	$Y_{1n_1} = (Y_{1n_1 1} Y_{1n_1 2} Y_{1n_1 3} Y_{1n_1 4} Y_{1n_1 5} \ Y_{1n_1 6} Y_{1n_1 7} Y_{1n_1 8} Y_{1n_1 9})$												
<hr/>													
	$C_2 Y_{21} = (Y_{211} Y_{212} Y_{213} Y_{214} Y_{215} \ Y_{216} Y_{217} Y_{218} Y_{219})$												
	$Y_{12} = (Y_{221} Y_{222} Y_{223} Y_{224} Y_{225} \ Y_{226} Y_{227} Y_{228} Y_{229})$												
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$						
$\vdots$	$\vdots$												
	$Y_{2n_1} = (Y_{2n_1 1} Y_{2n_1 2} Y_{2n_1 3} Y_{2n_1 4} Y_{2n_1 5} \ Y_{2n_1 6} Y_{2n_1 7} Y_{2n_1 8} Y_{2n_1 9})$												
<hr/>													
	$C_1 Y_{31} = (Y_{311} Y_{312} Y_{313} Y_{314} Y_{315} \ Y_{316} Y_{317} Y_{318} Y_{319})$												
	$Y_{32} = (Y_{321} Y_{322} Y_{323} Y_{324} Y_{325} \ Y_{326} Y_{327} Y_{328} Y_{329})$												
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$						
$\vdots$	$\vdots$												
	$Y_{3n_1} = (Y_{3n_1 1} Y_{3n_1 2} Y_{3n_1 3} Y_{3n_1 4} Y_{3n_1 5} \ Y_{3n_1 6} Y_{3n_1 7} Y_{3n_1 8} Y_{3n_1 9})$												

Para verificar la hipótesis sobre el factor A, el factor B y la interacción AB, se emplean contrastes de hipótesis<sup>2</sup> entre  $Y_{ij}$ . Algunos de estos contrastes, se presentan a través de las siguientes matrices:

$$A_1B_1A_1B_2A_1B_3A_2B_1A_2B_2A_2B_3A_3B_1A_3B_2A_3B_3$$

$$A = \begin{pmatrix} 2 & 2 & 2 & -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 \end{pmatrix} \quad (45)$$

$$A_1B_1A_1B_2A_1B_3A_2B_1A_2B_2A_2B_3A_3B_1A_3B_2A_3B_3$$

$$B = \begin{pmatrix} 2 & -1 & -12 & -1 & -1 & 2 & -1 & -1 \\ 0 & 1 & -10 & 1 & -10 & 1 & -1 & -1 \end{pmatrix} \quad (46)$$

$$A_1B_1A_1B_2A_1B_3A_2B_1A_2B_2A_2B_3A_3B_1A_3B_2A_3B_3 \quad (47)$$

$$C = \begin{pmatrix} 4 & -2 & -2 & -2 & 1 & 1 & -2 & 11 \\ 0 & 2 & -2 & 0 & -1 & 1 & 0 & -11 \\ 0 & 0 & 0 & -2 & -1 & -1 & -2 & 11 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -11 \end{pmatrix} \quad (48)$$

Las filas de la matriz  $A$  corresponden a contrastes ortogonales entre los niveles del factor A, los cuales comparan los siguientes niveles:

- a. El nivel  $A_1$  frente a los niveles  $A_2$  y  $A_3$  conjuntamente, y
- b. El nivel  $A_2$  frente al nivel  $A_3$ .

En forma semejante, las filas de la matriz  $B$  contienen los contrastes:

- a. El nivel  $B_1$  frente a los niveles  $B_2$  y  $B_3$  conjuntamente, y
- b. El nivel  $B_2$  frente al nivel  $B_3$ .

---

<sup>2</sup>Proceso mediante el cual se intenta comprobar si una afirmación sobre alguna propiedad poblacional puede ser sostenida a la luz de la información muestral disponible (Pardo y San Martín, 1998)

Cabe aclarar que es posible construir otros contrastes ortogonales para el factor  $A$  y el factor  $B$ . La matriz  $C$  está asociada con las interacciones entre los dos factores, y se obtiene como el producto entre los respectivos elementos de las filas de la matriz  $A$  y los de las filas de  $B$ .

Como podemos observar en el caso anterior se calcula por medio de la siguiente formula:

$$Y = \sum_{ij} \frac{Y_{ij}}{N}, S_p = \frac{E}{(N-q)}, N = \sum_i n_i. (49)$$

Si la matriz  $C$  tiene  $q$  niveles con medias  $\mu_1 \dots \mu_q$ . Entonces se verifican con la siguiente estadística

$$T_A^2 = N (AY..)' (A' S_p A)^{-1} (AY..) (50)$$

La cual se distribuye como:  $T_{(2, N-q)}^2$ , aquí el 2 que va en el subíndice, corresponde al número de filas de la matriz  $A$ .

Las hipótesis  $H_0: B\mu = 0$  y  $H_0: C\mu = 0$ , para los efectos principales de  $B$  y las interacciones entre  $A$  y  $B$ , se verifican de manera similar con las estadísticas, respectivamente:

$$T_B^2 = N (BY..)' (B' S_p B)^{-1} (BY..) (51)$$

$$T_{AB}^2 = N (PY..)' (P' S_p P)^{-1} (PY..) (52)$$

Las cuales se distribuyen como  $T_{(2, N-q)}^2$  y  $T_{(4, N-q)}^2$ , respectivamente. En general, si el factor  $A$  tiene  $a$  niveles y el factor  $B$  tiene  $b$  niveles, entonces las matrices de contrastes  $A$ ,  $B$  y  $C$  tienen  $(a - 1)$ ,  $(b - 1)$  y  $(a - 1)(b - 1)$  filas, respectivamente.

Las estadísticas de prueba se distribuyen, en general y respectivamente como:

$$T^2_{(a-1, N-q)}, T^2_{(b-1, N-q)} \text{ y } T^2_{((a-1)(b-1), N-q)}, (53)$$

Una prueba alternativa, para los efectos principales  $A$  y  $B$  y la interacción entre éstos es la lambda de Wilks ( $\Lambda$ ) en la cual se particiona “la suma de cuadrados total” como:

$$\sum_{ij} Y_{ij} Y'_{ij} = E + (H + H^*), (54)$$

Donde,

$H^* = N\bar{Y}.. \bar{Y}'$ . La hipótesis de interés es  $H_{0A}: A\bar{\mu} = 0$ , la cual se contrasta mediante la estadística

$$\Lambda_A = \frac{|AEA'|}{|A(E+H^*)A'|} (55)$$

La cual, bajo  $H_0$ , se distribuye como  $\Lambda_{(\alpha-1, 1, N_q)}$ , con  $\alpha$  el número de niveles del Factor  $A$ .

La dimensión de esta es  $(a-1)$  porque la matriz  $AEA'$  es de tamaño  $((a-1) * (a-1))$ .

Estadísticas similares para verificar los efectos del factor  $B$  y las interacciones entre  $A$  y  $B$ .

Los efectos principales del factor  $C$ , como en el caso de *medidas repetidas en q muestras*, son equivalentes a verificar la hipótesis

$$H_{C_0}: 1'\mu_1 = \dots = 1'\mu_q, (56)$$

Al igual que la hipótesis planteada en la igualdad  $H_0: C\mu_1 = \dots = C\mu_q$ , esta verificada con una estadística  $F$  univariado sobre los  $Z_{ij} = 1'Y_{ij}$ , en la forma de un análisis de varianza univariado a una vía de clasificación.

Las interacciones tipo  $AC, BC$  y  $ABC$  se prueban en la forma siguiente:

- Interacción  $AC$ . Equivale a la hipótesis

$$H_{AC_0}: A\mu_1 = \dots = A\mu_q (57)$$

La cual establece que los contrastes en el factor  $A$  son los mismos a través de todos los  $q$  niveles del factor  $C$ . Una estadística para verificar esta hipótesis es

$$\Lambda_{AC} = \frac{|AEA'|}{|A(E+H)A'|} \quad (58)$$

La cual se distribuye como  $\Lambda_{(\alpha-1, q-1, N_q)}$ . La hipótesis anterior se puede contrastar a través de un análisis de varianza multivariado para un modelo a una vía de clasificación, sobre los vectores de observaciones transformados a  $Z_{ij} = AY_{ij}$ .

- Interacción  $BC$ , Se expresa a través de la hipótesis

$$H_{BC_0}: B\mu_1 = \dots = B\mu_q \quad (59)$$

La cual se verifica a través de la estadística

$$\Lambda_{BC} = \frac{|BEB'|}{|B(E+H)B'|} \quad (60)$$

La cual se distribuye como  $\Lambda_{(b-1, q-1, N_q)}$ . También se puede verificar con el análisis de varianza multivariado sobre los  $Z_{ij} = BY_{ij}$ .

- Interacción  $ABC$ . Se expresa mediante la hipótesis

$$H_{ABC_0}: P\mu_1 = \dots = P\mu_q \quad (61)$$

La cual se contrasta mediante la estadística

$$\Lambda_{BC} = \frac{|PEP'|}{|P(E+H)BP'|} \quad (62)$$

La cual se distribuye como  $\Lambda_{((a-1)(b-1), q-1, N-q)}$ . También se puede verificar con el análisis de varianza multivariado sobre los  $Z_{ij} = PY_{ij}$ .

Las pruebas sobre los contrastes AC, BC o ABC se pueden desarrollar a través de los valores propios de las matrices asociadas a “covariación entre” y al “covariación dentro”. Así por ejemplo, para la interacción tipo AC se obtienen los valores propios de la matriz  $(AEA')^{-1}(AHA')$ , y con ellos se calculan estadísticas como la traza de Lawley-Hotelling, la traza de Barlett-Nanda-Pillai o el máximo valor propio de Roy.

## 5.6. MODELOS MIXTOS

Este tipo de modelos recibe distintos nombres que destacan una u otra de sus principales características. Se denominan modelos multinivel o jerárquicos, porque tienen en cuenta la estructura agregada de los datos en distintos niveles (clases) que pueden estar ordenados o no jerárquicamente (Seoane, J. 2004).

Se denominan también mixtos o de efectos mixtos (“mixed models” o “mixed-effects models”) porque combinan efectos fijos y aleatorios. Aclaremos estos aspectos. Los modelos mixtos siguen una estrategia lógica propia de muchos otros tipos de modelos estadísticos por la que se trata de describir la relación entre una “variable respuesta” (“dependiente”) y una o varias “variables explicativas” (alias “independientes”, “predictores” o “covariables”).

Por ejemplo, la relación entre los procesos de los ecosistemas y la diversidad funcional se puede medir en conjuntos de parcelas de estudio localizados en distintos países, el cambio en la distribución de especies montanas se puede observar en cumbres que se agregan dentro de diferentes macizos o se examinan en distintos periodos temporales, el efecto de la variación en las estrategias reproductoras puede examinarse en la descendencia de un territorio, nido o pareja y el comportamiento de una especie se puede describir a través de medidas realizadas en poblaciones (utilizando individuos) o en individuos (empleando muestras de los mismos sujetos obtenidas en diferentes ocasiones). (Seoane, J, 2004).

En todos estos casos las respuestas (procesos, distribución, descendencia y comportamiento) se estudian mediante datos que se agregan en unidades temporales o espaciales (países, montañas y años, nidos, individuos y estaciones) y que, a su vez, pueden estar agregadas jerárquicamente en niveles (individuos dentro del nivel superior de estaciones).

Un detalle crucial de estos datos es que a menudo las medidas realizadas dentro de un mismo nivel de agregación no son independientes y así, por ejemplo, cabe esperar que las medidas hechas a un mismo individuo estén relacionadas entre sí; lo que incurre en pseudoreplicación y, en consecuencia, en el problema que se conoce como Error de Tipo I (rechazar la hipótesis nula cuando esta es cierta).

Los modelos mixtos permiten acomodar esta situación incorporando al menos una variable explicativa de tipo categórico que represente esas unidades en que se agregan los datos. Los valores que se dan a esa variable son clases o categorías identificativas sin valor numérico (individuos “i1”, “i2”, etc). Los efectos, es decir la influencia, de las variables explicativas sobre la respuesta se miden a través de distintos parámetros de los modelos.

Si los valores que puede tener la variable explicativa son informativos y su número está fijado de antemano estamos ante lo que se denomina “**efectos fijos**”, que asumen que tales valores son independientes entre observaciones. Si los valores de la variable son solo identificativos y podrían encontrarse otros si el estudio se repitiera en diferentes circunstancias (al año siguiente o en un lugar distinto) estamos ante “**efectos aleatorios**”, en los que se asume una relación entre las observaciones realizadas en la misma clase.

Si la variable es de tipo categórico y adquiere un bajo número de valores distintos o “niveles” (1 a 4) es mejor tratarla como de efecto fijo y entender que nuestros resultados no pueden extrapolarse más allá de esos niveles.

Por el contrario, si la variable adquiere un alto número de valores ( $> 4-5$  pero preferiblemente  $> 10$ ) es preferible considerarla de efecto aleatorio, lo que permite generalizar los resultados al universo del que nuestros datos son una muestra. Un ejemplo típico de variable de efecto fijo es el sexo, pues su conjunto posible de valores está limitado (macho y hembra), estos son informativos (el sexo acarrea consecuencias fisiológicas y comportamentales) y la medida de la respuesta se asume independiente entre observaciones (entre los distintos machos).

En contraste, el territorio de reproducción de un animal se suele considerar como una variable de efecto aleatorio, pues aquellos que se muestrearon son normalmente una submuestra aleatoria del conjunto de la población y la respuesta a analizar se espera que esté correlacionada dentro de cada uno (la inversión parental que se mide en el mismo territorio en distintos momentos temporales). En los modelos mixtos, los efectos fijos se ajustan mediante parámetros como el intercepto y la pendiente en una regresión, mientras que los efectos aleatorios son, estrictamente, variables aleatorias que no se observan, pero cuya distribución puede estimarse mediante la varianza de una distribución normal.

Los modelos mixtos permiten modelar la respuesta de un estudio experimental u observacional como función de factores o covariables cuyos efectos pueden considerarse tanto como fijos o aleatorios. Cada modelo estadístico que contiene una media general  $\mu$ , es un modelo mixto por definición, ya que también contiene un término de error aleatorio, y por tanto contiene ambos tipos de efectos. Sin embargo, en la práctica, el nombre modelo mixto se reserva usualmente para cualquier modelo que contiene efectos fijos distintos a  $\mu$  y efectos aleatorios diferentes a los errores aleatorios. (Balzarini *et al.* 2004).

Además, Balzarini *et al.* (2004), dice que los supuestos clásicos de independencia y homogeneidad de varianzas para los términos aleatorios del modelo lineal general, se flexibilizan en el marco del modelo mixto. Tanto la estructura de correlaciones como la



presencia de varianzas heterogéneas pueden ser especificadas a través de la modelación pues esta característica hace a los modelos mixtos muy interesantes.

Los modelos mixtos se adecúan bien en diferentes situaciones, como en la agricultura y forestaría, como por ejemplo cuando existe algún tipo de estructura de bloqueo de unidades experimentales que afecta las covarianzas entre observaciones. Ilustran este tipo de situación aquellos estudios donde el material experimental se evalúa en varios ensayos y por tanto es razonable asumir que existen correlaciones entre observaciones del mismo ensayo. La modelación en el marco de los modelos mixtos maneja estas correlaciones mediante la incorporación de variables aleatorias o mediante la modelación directa de la matriz de covarianzas residual. Existen muchos beneficios que pueden obtenerse con el uso de modelos mixtos. En algunas situaciones se incrementa la precisión de las estimaciones. En otras se contempla mejor la estructura y se amplía el espacio de inferencia, sobre todo cuando la estructura de los datos es jerárquica.

#### 5.6.1. Ventajas

- Brindan la posibilidad de analizar datos con estructuras de dependencia, desbalances y varianzas heterogéneas.
- Permite contemplar la falta de cumplimiento de los supuestos tradicionales y modelar, de manera flexible, complicadas estructuras de datos.
- Se incrementa la precisión de las estimaciones.
- Se amplía el espacio de inferencia y se comprende mejor la estructura de datos.
- Los modelos mixtos permiten la estimación eficiente del vector de parámetros fijos  $\hat{\beta}$  que definen el patrón común de la población.
- Predice de forma eficiente del vector de parámetros aleatorios  $\hat{b}_i$  específico para la unidad de muestreo  $i$ , que define el patrón común de desviación de los componentes de la unidad con respecto de la media.
- Descomposición de la varianza total en la varianza entre las unidades de muestreo y la varianza dentro de cada unidad. Estimación de los componentes de la varianza que definen las matrices de varianza estimadas  $\hat{D}$  y  $\hat{R}$ .

- Calibración del modelo para una unidad  $K$  no muestreada, al poder predecir el vector de parámetros aleatorios  $\widehat{b}_k$  si se dispone de al menos una medición de la variable de interés en la nueva unidad.

El procedimiento de los Modelos lineales mixtos amplía el modelo lineal general de manera que los datos puedan presentar variabilidad correlacionada y no constante. El modelo lineal mixto proporciona, por tanto, la flexibilidad necesaria para modelar no sólo las medias sino también las varianzas y covarianzas de los datos.

El procedimiento Modelos lineales mixtos es asimismo una herramienta flexible para ajustar otros modelos que puedan ser formulados como modelos lineales mixtos. Dichos modelos incluyen los modelos multinivel, los modelos lineales jerárquicos y los modelos con coeficientes aleatorios.

El objetivo al implementar un análisis basado en modelos lineales mixtos es la estimación de los  $\beta'$  y la predicción de los valores de los  $b'_i$ s, además de estimar las componentes de varianza.

El modelo dado por la expresión:

$$Y = X\beta + Zb + \varepsilon, (63)$$

Donde

$y$ : Es el vector  $n \times 1 [y_1^t, \dots, y_n^t]^t$  que contiene las observaciones medidas.

$x$ : Es la matriz  $n \times p$ , cuyas filas son los vectores  $x_{ij}$ .

$\beta$  : Es el vector de parámetros fijos definido.

$Z$ : Es la matriz en bloque diagonal cuyos bloques son las  $r$  matrices  $Z_i$ .

$b$  : Es el vector  $[b_1^t, b_2^t, \dots, b_r^t]^t$  que incluye los  $q$  – parámetros aleatorios para cada una de las  $r$  parcelas, tal que  $b \sim N(0, D)$ .

$\varepsilon$  : Es un vector  $n \times 1$  que contiene los términos residuales del error, tal que  $\varepsilon \sim N(0, R)$ .

En el caso más común  $D$  es la matriz en bloque diagonal compuesta por las  $r$  matrices  $D_i (D_1 = \dots = D_i = \dots = D_r)$  y  $R$  es una matriz diagonal  $n \times n$ , cuya única componente es la varianza residual del modelo  $\sigma_e^2$ . [60]

El modelo mostrado en la ecuación (60) puede expresarse de manera más detallada, así:

$$\begin{bmatrix} y_{i1} \\ \vdots \\ y_{ij} \\ \vdots \\ y_{ini} \end{bmatrix} = \begin{bmatrix} 1 & X_{i11} & X_{i1k} \\ 1 & X_{i21} & X_{i2k} \\ \vdots & \vdots & \vdots \\ 1 & X_{ij1} & X_{ijk} \\ \vdots & \vdots & \vdots \\ 1 & \dots & X_{inik} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} 1 & Z_{i11} & X_{i1r-1} \\ 1 & Z_{i21} & X_{i2r-1} \\ \vdots & \vdots & \vdots \\ 1 & Z_{ij1} & X_{ijk-1} \\ \vdots & \vdots & \vdots \\ 1 & \dots & X_{inik-1} \end{bmatrix} \begin{bmatrix} b_{i0} \\ b_{i1} \\ \vdots \\ b_{ij} \\ \vdots \\ b_{ir-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{ij} \\ \vdots \\ \varepsilon_{ini} \end{bmatrix} \quad (64)$$

Donde el  $i$  – ésimo sujeto de  $y$  tiene  $n_i$  observaciones:  $(y_{i1}, y_{i2}, \dots, y_{ini})$  con:

$$i = 1, \dots, N, N \text{ Es el número de sujetos.}$$

La matriz  $Z$  puede expresarse como una matriz de bloques conformada por las submatrices  $Z_i$ , con  $i = 1, \dots, N$ .

$$\mathbf{Z}_{T \times Nr} = \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Z_N \end{bmatrix} \quad (65)$$

Aquí, cada  $Z_i$  tiene dimensión  $n_i \times r$ . Si el modelo es balanceado definiendo:

$$n_i = k, \text{ se tiene que } T = N_k, \text{ así el orden de } \mathbf{Z} \text{ será } N_k \times Nr$$

Las matrices diagonales de varianza-covarianza  $R$  y  $B$ , del error y del efecto aleatorio respectivamente, pueden expresarse así:

$$R_{T \times T} = \begin{bmatrix} R_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_N \end{bmatrix} \mathbf{B}_{Nr \times Nr} = \begin{bmatrix} D_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & D_N \end{bmatrix} \quad (66)$$

Donde  $R_i$  es la matriz de orden  $n_i \times n_i$ , y si el diseño es balanceado podemos decir que:  $n_i = k$ , luego, la matriz  $R_i$  sería de orden  $k \times k$ . La matriz  $D_i$  es de orden  $r \times r$ . Usualmente la estructura de varianza-covarianza se define a partir de la matriz  $D_i$ .

Como ejemplo particular se considera al intercepto como aleatorio, puede expresarse así:

$$Y_{ij} = (\beta_0 + b_i) + x_{ij}\beta_1 + \varepsilon_{ij}, (67)$$

Donde,

$Y_{ij}$  Es la respuesta del sujeto  $i$  en la medida  $j$ , con  $i = 1, \dots, n$ . y  $j = 1, \dots, n_i$ .  $X_{ij}$  Representa la covariable de la matriz de diseño  $X$  de los efectos fijos para el sujeto  $i$  en la medida  $j$ , con  $i = 1, \dots, N$  y  $j = 1, \dots, n_i$ , los valores  $\beta_0$  y  $\beta_1$  son los efectos fijos del intercepto y la covariable  $x_{ij}$  respectivamente,  $b_i$  representa el intercepto aleatorio en el modelo y  $\varepsilon_{ij}$  Es el término de error para el sujeto  $i$  en la medida  $j$ .

**5.6.2. Estructura de Covarianza.** El modelo más simple es el modelo de covarianza independiente, donde el error está dentro del sujeto. La correlación es cero, y por tanto  $\Sigma = \sigma^2 I$ . El más complejo es el modelo de covarianza no estructurada, donde los errores están dentro del sujeto para cada par de veces tienen su propia correlación única. Así

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1k} \\ & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2k} \\ & & & \cdots & \sigma_{3k} \\ & & & & \vdots \\ & & & & \sigma_k^2 \end{bmatrix} (68)$$

En algunas aplicaciones, la correlación dentro del sujeto es mínima o nula. Por ejemplo, en algunos ensayos de nutrición agronómica y de animales grandes, las mediciones repetidas pueden ocurrir a intervalos suficientemente largos, tales como mensuales, que la correlación es efectivamente cero en relación con otra variación. En tales casos, la estructura de independencia es aceptable. Sin embargo, esto debe comprobarse antes de analizar los datos suponiendo errores no correlacionados.

La correlación está presente en la mayoría de los datos de medidas repetidas hasta cierto punto. Sin embargo, la correlación no suele ser tan compleja como el modelo no estructurado. El modelo más simple con correlación es la simetría compuesta, denominada (CS)

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & P & P & \dots & P \\ & 1 & P & \dots & P \\ & & 1 & \dots & P \\ & & & \ddots & P \\ & & & & 1 \end{bmatrix} \quad (69)$$

Suponemos que la correlación es constante independientemente del desfase entre pares de mediciones repetidas. Obsérvese que la ecuación del modelo de parcela dividida

$$Y_{ijk} = \mu + \alpha_i + b_{ij} + \gamma_k + (\alpha\gamma)_{ik} + e_{ijk} \quad (70)$$

Puede expresarse como

$$Y_{ijk} = \mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} + \varepsilon_{ijk} \quad (71)$$

Suponiendo que  $\varepsilon_{ijk} = b_{ij} + e_{ijk}$  distribuido  $N(0, \sigma^2)$  y que el  $b_{ij}$  se distribuyen  $N(0, \sigma^2)$  induce la covarianza de simetría compuesta de la  $\varepsilon_{ijk}$  para el  $ij$  –esimo sujeto. Más específicamente,  $Var[Y_{ijk}] = \sigma_B^2 + \sigma_s^2$  y  $Cov[Y_{ijk}, Y_{ijk'}] = \sigma_B^2$  para  $K \neq K'$ , idéntico al modelo de simetría compuesta (CS) con  $\sigma^2 = \sigma_B^2 + \sigma_s^2$  y  $P = \sigma_B^2 / (\sigma_s^2 + \sigma_B^2)$ .

Típicamente, la correlación entre las observaciones es una función de su retraso en el tiempo: las observaciones adyacentes tienden a estar más altamente correlacionadas que las observaciones más distanciadas en el tiempo. Varios modelos pueden describir adecuadamente en esta correlación. Quizás el más comúnmente utilizado es el modelo de primer orden auto regresivo, o AR (1). Para el modelo AR (1)

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & P & P^2 & \dots & P^{K-1} \\ & 1 & P & \dots & P^{K-2} \\ & & 1 & \dots & P \\ & & & \ddots & P \\ & & & & 1 \end{bmatrix} \quad (72)$$

El modelo AR (1) asume que  $e_{ijk} = Pe_{ijk-1} + S_{ijk}$ , donde  $S_{ijk} \sim N(0, \sigma_s^2)$ . Se deduce que  $\sigma^2 = \frac{\sigma_s^2}{1-P^2}$ . Esto ayuda a explicar por qué los modelos de error independientes tienden a subestimarse con la varianza del sujeto cuando la correlación entre los errores es no despreciable.

Bajo el modelo AR (1), la correlación entre los errores dentro del sujeto adyacentes es  $\rho$ , Independientemente de si el par de observaciones es el primero y segundo, segundo y tercero, o  $(K - 1)$  veces  $K$  veces, mientras que, con el modelo no estructurado, cada par tiene su propia correlación. La correlación es  $P^2$  para cualquier par de errores 2 unidades aparte, como el primero y el tercero. En general, los errores  $d$  unidades aparte tienen correlación  $P^d$ . Obsérvese que el modelo AR (1) requiere estimaciones de sólo dos parámetros,  $\sigma^2$  y  $P^2$ , mientras que los modelos no estructurados requieren la estimación de los parámetros  $K + K(K - 1)/2$ .

El modelo Toeplitz es similar al modelo AR (1) en el sentido de que pares de errores dentro del sujeto separados por un retardo común comparten la misma correlación. Sin embargo, los errores  $d$  unidades aparte tienen correlación  $P_d$  en lugar de  $P^d$ . Así, para el modelo Toeplitz,

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & P & P^2 & \dots & P^{K-1} \\ & 1 & P & \dots & P^{K-2} \\ & & 1 & \dots & P \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix} = \begin{bmatrix} \sigma_0^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1,k} \\ & \sigma_0^2 & \sigma_{21} & \dots & \sigma_{2,k} \\ & & \sigma_0^2 & \dots & \vdots \\ & & & \ddots & \sigma_{k-1,k} \\ & & & & \sigma_0^2 \end{bmatrix} \quad (73)$$

El modelo de Toeplitz es menos restrictivo que el modelo AR (1), pero requiere parámetros  $K$  ( $\sigma^2, P_1, \dots, P_{K-1}$ ) en lugar de sólo dos.

Los modelos AR (1) y Toeplitz tienen sentido cuando las observaciones están igualmente espaciadas y la estructura de correlación no cambia apreciablemente con el tiempo. Un modelo más general que preserve las características principales de estos modelos, pero

que permite un espaciamiento y un cambio desiguales a lo largo del tiempo, es el modelo de ante-dependencia de primer orden, o ANTE (1). La estructura del modelo es

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2P_1 & \sigma_1\sigma_3P_1P_2 & \dots & \sigma_1\sigma_2P_1P_2 \dots P_{k-1} \\ & \sigma_2^2 & \sigma_2\sigma_3P_2 & \dots & \sigma_2\sigma_3P_2P_3 \dots P_{k-1} \\ & & \sigma_3^2 & \dots & \vdots \\ & & & \ddots & \sigma_{k-1}\sigma_kP_{k-1} \\ & & & & \sigma_k^2 \end{bmatrix} \quad (74)$$

Puede verse que el modelo ANTE (1) asume que la varianza entre las observaciones cambia con el tiempo y que la correlación entre pares de observaciones es el producto de las correlaciones entre momentos adyacentes de las observaciones, de modo que la correlación puede cambiar con el tiempo. El modelo ANTE (1) requiere estimar parámetros  $2K - 1$ .

Otras estructuras pueden ser modificadas para acomodar variaciones heterogéneas en el tiempo, incluyendo el modelo autor regresivo de primer orden y Toeplitz.

Todavía otra covarianza se deriva de combinar CS con AR (1). El componente CS modela la variación entre las medias del sujeto y el componente AR (1) modela el componente dentro del sujeto, condicional a un sujeto en particular, que incluye el efecto aleatorio para el sujeto, con la estructura AR (1) impuesta a  $R$ . Por lo tanto, se puede llamar AR (1) + RE. Tiene la forma matricial

$$\Sigma = \sigma_B^2 \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ & 1 & 1 & \dots & 1 \\ & & 1 & \dots & 1 \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix} + \sigma_S^2 \begin{bmatrix} 1 & P & P^2 & \dots & P^{K-1} \\ & 1 & P & \dots & P^{K-2} \\ & & 1 & \dots & P \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix} \quad (75)$$

Donde  $\sigma_B^2$  Es la varianza entre sujetos y  $\sigma_S^2$  Es la variación condicional de un sujeto.

## 5.7. CRITERIO DE INFORMACIÓN AKAIKE (AIC)

En la derivación del AIC por Akaike (1974), de entrada, se considera la situación donde  $g(x) = f(x|\theta_0)$ , es decir, la densidad de probabilidad  $g(x)$  verdadera se encuentra

incluida en la familia dada,  $\{f(x|\theta): \theta \in \Theta \subset R^p\}$ . Si  $K(\theta_0, \theta)$  denota  $I(g, f(\cdot|\theta))$  y además si  $\theta$  esta suficientemente cercano a  $\theta_0$ ,  $K(\theta_0, \theta)$  se puede aproximar por

$$K(\theta_0, \theta_0 + \Delta\theta) \approx \frac{1}{2} \Delta\theta' I(\theta_0) \Delta\theta \quad (76)$$

Donde,

$$I(\theta_0) = \int g(x) \frac{\delta \log[f(x|\theta_0)]}{\delta \theta} \frac{\delta \log[f(x|\theta_0)]}{\delta \theta} dx \quad (77)$$

$$\text{Y} \quad \frac{\delta \log[f(x|\theta_0)]}{\delta \theta} := \frac{\delta \log[f(x|\theta)]}{\delta \theta} \quad (78)$$

Con lo anterior se llega a que

$$AIC = -2l(\widehat{\theta_{n,k}}) + 2k \quad (79)$$

El cual es el criterio de Akaike

## 5.8. CRITERIO DE INFORMACIÓN BAYESIANA (BIC)

El criterio de información Bayesiana (BIC) propuesto por Schwartz en (1978), ha sido uno de los métodos más populares usado para la selección de modelos. Este es un criterio de evaluación de modelos en términos de sus probabilidades posteriores.

Se denota

$$BIC =: -2l_{n,i}(\widehat{\theta_n^2}) + k_i \log(n) \quad (80)$$



## **6. APLICACIONES**

### **6.1. ASPECTOS GENERALES**

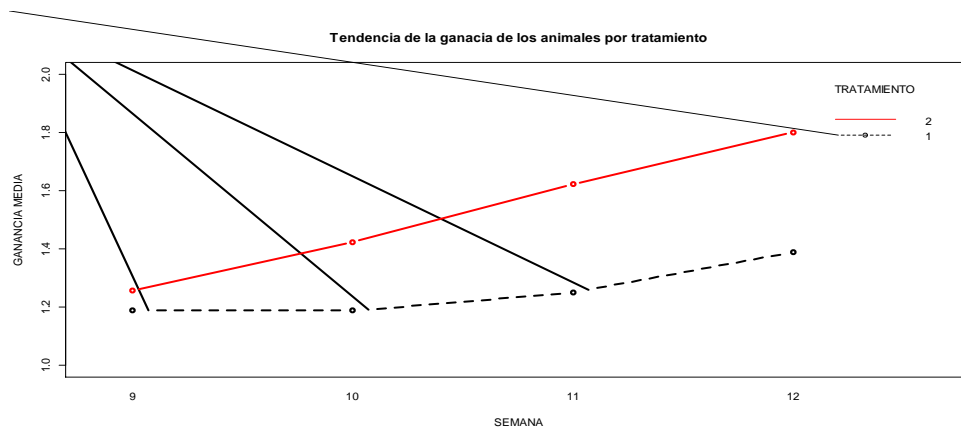
En la ganadería es muy común realizar ensayos en los que se dispone de varias observaciones del mismo individuo y se pueden tener también varios grupos; una de estas situaciones se ve reflejada en la ganancia de peso de reces jóvenes, la cual se mide del mismo animal repetidas veces, según el tiempo fijado y a ello se obtiene varias mediciones de un mismo sujeto. A este tipo de estudio se le denomina medidas repetidas y se utiliza en situaciones en las que la variable respuesta en cada unidad experimental se mide en múltiples ocasiones y probablemente en condiciones experimentales.

Para este tipo de casos con diseños de medidas repetidas se realizará diversos métodos estadísticos que se usan frecuentemente en estas situaciones, exponiendo sus desventajas y ventajas ante las medidas repetidas; para ello se expondrá un análisis detallado usando los datos que se encuentran en (Miroslav Kaps y William R. Lamberson, 2004). Los datos obtenidos son un estudio realizado con reces jóvenes, en el cual miden su ganancia de peso aplicando dos tratamientos diferentes a dos grupos de reces en un periodo de tiempo de 9 a 12 semanas, obteniendo así los datos esperados para un análisis de medidas repetidas.

Fueron seleccionados 17 animales separados en dos grupos según el tratamiento que se le aplicó, es decir, el tratamiento “1” fue aplicado a los animales que van del 1 al 8, y el tratamiento “2” fue aplicado a los animales que van del 9 al 17; estos tratamientos fueron aplicados en cuatro semanas seguidas “semana 9, semana 10, semana 11 y semana 12”, obteniendo así su ganancia de peso obtenida por el tratamiento dado en cuatro semanas, logrando así medidas a través del tiempo. Se realizaron tres tipos de análisis para luego comparar cuál de ellos es el mejor, exponiendo sus ventajas y debilidades ante este tipo de datos. Este estudio se realizó usando el software estadístico R-Project y.

A continuación, se mostrará el gráfico de perfiles, exponiendo el comportamiento de los datos.

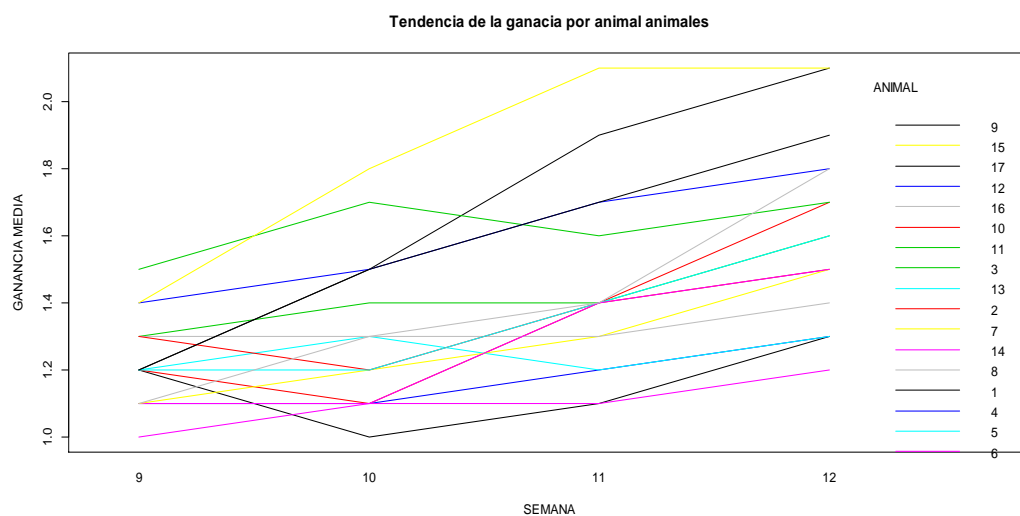
**Figura 2.** Tendencia de ganancia de peso por semana de los tratamientos 1 y 2.



Fuente: R-Project (2017)

Como se observa en la Figura anterior, vemos que el tratamiento “2” obtuvo más ganancia a comparación del tratamiento “1”, a través de las semanas 9, 10, 11 y 12. Ahora vamos a realizar la Figura de ganancia de peso por animal.

**Figura. 3.** Tendencia de ganancia de peso por semana de los 17 animales.



Fuente: R-Project (2017)

Como se observa existe gran variabilidad entre los animales, además se ve que los animales con mayor ganancia de peso corresponden en su mayoría al tratamiento 2 y los de menor ganancia corresponden al tratamiento 1.

Para obtener información general de los datos se realizó un análisis descriptivo por semana (tabla 1) y por tratamiento (tabla)

**Tabla 4.** Estadística descriptiva por semana.

<b>Semana</b>	<b>Media</b>	<b>Mediana</b>	<b>Varianza</b>	<b>Desviación Estándar</b>
<b>9</b>	1,22	1,2	0,016	0,13
<b>10</b>	1,31	1,3	0,051	0,22
<b>11</b>	1,44	1,4	0,075	0,27
<b>12</b>	1,6	1,6	0,074	0,27

Como se observa a medida que pasan las semanas el peso de los animales va subiendo como lo refleja la media y mediana.

**Tabla 5.** Estadística descriptiva por tratamiento.

<b>Semana</b>	<b>Media</b>	<b>Mediana</b>	<b>Varianza</b>	<b>Desviación Estándar</b>
<b>1</b>	1,25	1,25	0,019	0,141
<b>2</b>	1,52	1,5	0,086	0,294

Como se observa el tratamiento 2 tuvo mejor ganancia de peso a comparación del tratamiento 1, además se mantiene la variabilidad entre tratamientos y el tratamiento 2 posee mayor variabilidad.

## 6.2. ANÁLISIS DE VARIANZA “ANOVA”

El modelo a analizar es el siguiente:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad (81)$$

Dónde:

$y_{ijk}$  : Es la variable respuesta (Ganancia de peso).

$\mu$  : Es la media poblacional.

$\alpha_i$  : Es el efecto semana en el  $i$ -ésimo nivel.

$\beta_j$  Es el efecto tratamiento en el  $j$ -ésimo nivel.

$(\alpha\beta)_{ij}$  : Es la interacción de los factores semana y tratamiento en el  $i, j$ -ésimo nivel.

$\varepsilon_{ijk}$  : Es el error que se distribuye normalmente con media 0 y varianza  $\sigma^2$ ,  $N(0, \sigma^2)$ .

Realizamos el ANOVA obteniendo los siguientes resultados,

**Tabla 6.** Anova

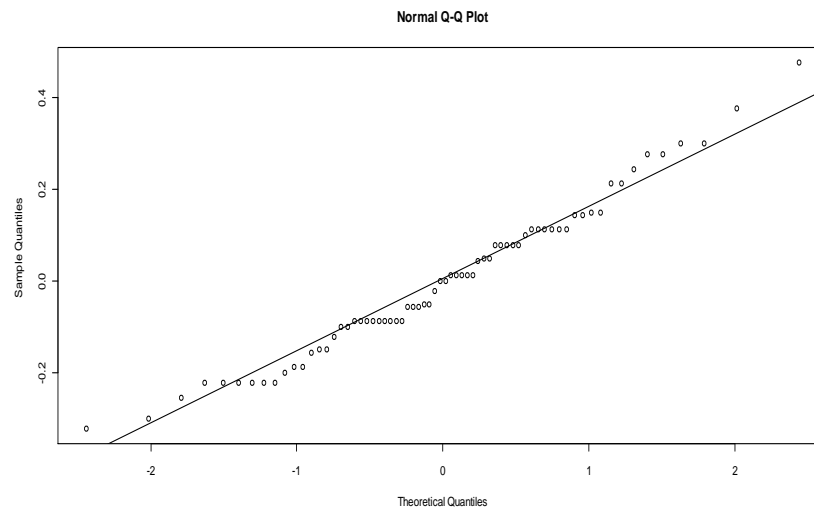
	GRADOS DE LIBERTAD	SUMA DE CUADRADOS	MEDIA	F - VALOR	P - VALOR
TRATAMIENTO	1	1,25	1,25	39,14	$4,56 \cdot 10^{-8}$
SEMANA	3	1,41	0,47	14,78	$2,55 \cdot 10^{-7}$
TRATAMIENTO*SEMANA	3	0,3	0,1	3,21	0,02

Como se puede observar en la interacción (TRATAMIENTO: SEMANA), hay efecto significativo de interacción entre los niveles de factor semana y tratamiento, ya que su P-valor es inferior al nivel de significancia escogido (0.05).

A continuación, se realiza el análisis de diagnóstico para ver qué tan bueno es el ajuste del modelo que se aplicó.

### 6.2.1 Prueba de Normalidad

**Figura. 4.** Figura de normalidad Q-Q Plot.



Fuente: R-Project (2017)

Como se logra ver en la **Figura**, se puede decir que hay normalidad, ya que los puntos se mantienen cercanos a la línea de normalidad. Ahora miremos el test de shapiro wilk.

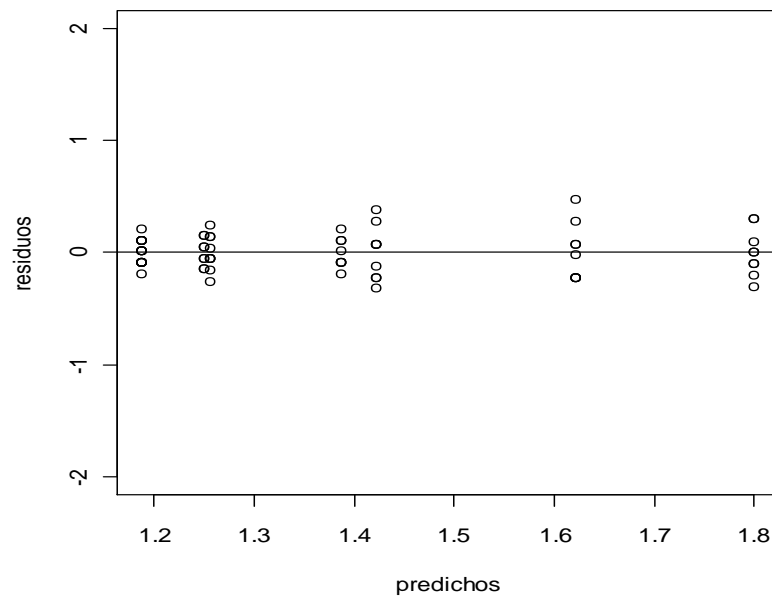
**Tabla 7.** Prueba7. Prueba de normalidad Shapiro Wilk.

Datos: residuos	Prueba de Normalidad
	Shapiro Wilk
W= 0,9774	P-Valor = 0,2541

Según el test de Shapiro Wilk de normalidad, se puede decir que hay normalidad, ya que el p-valor que arroja el test es superior a un nivel de significancia de 0.05, por lo tanto, se acepta la hipótesis nula de normalidad.

### 6.2.2 Prueba de Homogeneidad

**Figura 5.** Residuos vs predichos, para la Homoscedasticidad de varianzas.



Fuente: R-Project (2017)

Como se observa en la **Figura** es posible decir que no hay Homoscedasticidad de varianzas, pero no es viable asegurar que no hay Homoscedasticidad, por lo tanto se realizará un test de Levene.

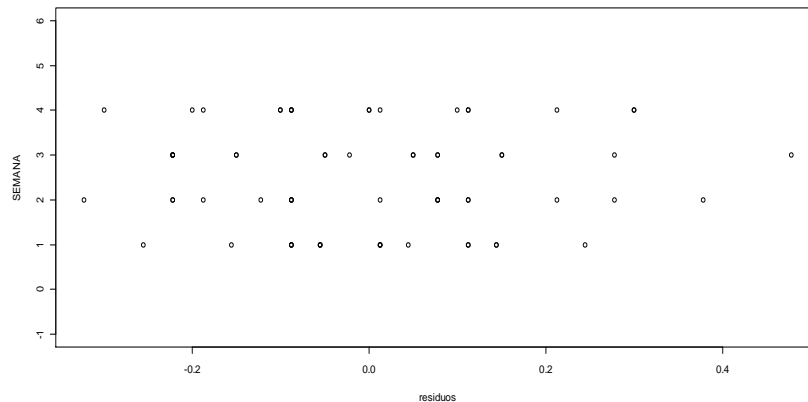
**Tabla 8.** Prueba de Levene para Homoscedasticidad de varianzas.

<b>Factor</b>	<b>Grados de Libertad</b>	<b>Suma de Cuadrados</b>	<b>Media de Cuadrados</b>	<b>F - Valor</b>	<b>P - Valor</b>
<b>Tratamiento</b>	1	0,091	0,091	10,58	0,0018
<b>Semana</b>	3	0,036	0,012	1,43	0,2422
<b>Tratamiento*S emana</b>	3	0,008	0,002	0,33	0,8033
<b>Residuales</b>	60	0,516	0,008		

Según el test de Levene si hay igualdad de varianzas en “semana (p-valor= 0.24)” y en la interacción “tratamiento: semana (p-valor=0.80)”, pero en “tratamiento (p-valor= 0.0018)” no hay igualdad de varianzas.

### 6.2.3 Prueba de Independencia

**Figura 6.** Figura residuos vs SEMANA, para la independencia.



Fuente: R-Proyect (2017)

En la Figura anterior se observa que no hay independencia, además vemos que los residuos poseen dependencia positiva, ya que existe correlación entre los datos.

En conclusión, el análisis de varianza no se ajusta bien para datos con medidas repetidas, igualmente cabe destacar que los datos son desbalanceados. Por lo tanto, vemos a ver si con el método de análisis multivariado se pueda ajustar a este tipo de datos.

### 6.3. ANÁLISIS MULTIVARIADO “MANOVA”

Para realizar este método estadístico los datos deben ser tomados como matrices, como se muestra en el modelo,

$$Y = \mu + X\beta + \varepsilon \quad (82)$$

$Y$  : Es la matriz  $n \times p$  de observaciones,  $n$  animales y  $p$  observaciones por animal a intervalos fijos del tiempo.

$\mu$ : Es el vector de media  $n \times 1$ .

$X\beta$  :  $X$  es el vector  $n \times 1$  de los tratamientos y  $\beta$  es la matriz  $n \times p$  de los parámetros de las semanas 9, 10, 11 y 12.

$\varepsilon$  : Es matriz  $n \times p$  de los errores, que se distribuyen normalmente  $N_p(0, \sigma_p)$ .

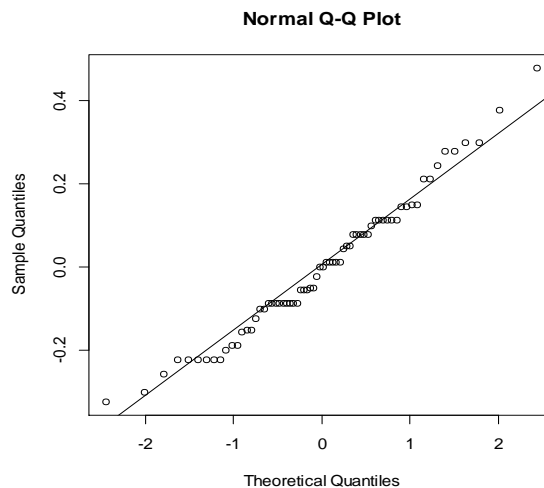
**Tabla 9. MANOVA**

	DF	PILLAI	APROXIMACION F	NUM F	DEN F	P - VALOR
<b>TRATAMIENTO</b>	1	0,6189	4,8723	4	12	0,0144
<b>RESIDUALES</b>	15					

Como se puede observar en el factor tratamiento hay diferencia significativa entre las medias, ya que el p-valor dado es inferior al nivel de significancia (0.05).

### 6.3.1 Prueba de Normalidad.

**Figura 7. Prueba de normalidad.**



Fuente: R-Project (2017)



Aparentemente se observa que hay normalidad en los datos, ahora se realizó una prueba de normalidad “SHAPIRO-WILK”.

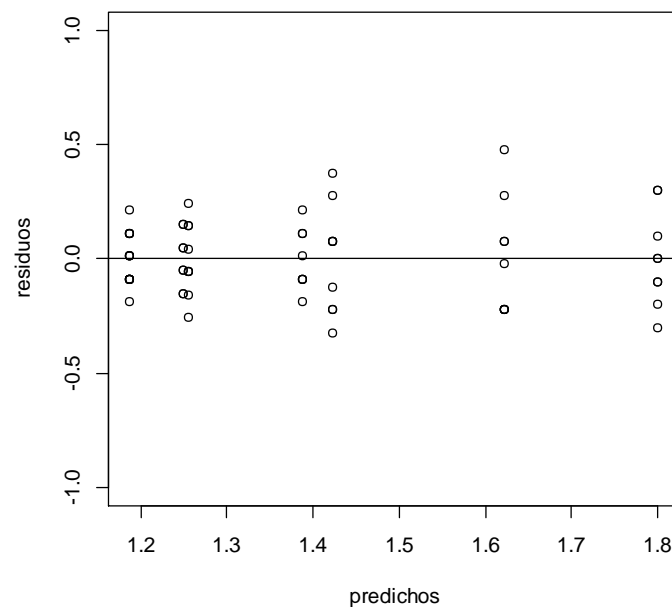
**Tabla 10** Prueba de normalidad SHAPIRO WILK.

PRUEBA SHAPIRO-WILK
DATOS: Residuos
W= 0,9774, P - Valor= 0,2541

Vemos que, si existe normalidad, ya que el p-valor es mayor que el nivel de significancia 0.05.

### 6.3.2 Prueba de Homoscedasticidad.

**Figura 8.** Grafica residuos vs predichos, para la Homoscedasticidad de varianzas.



Fuente: R- Proyect (2017)

Como se observa en la Figura anterior, puede ser que no exista Homoscedasticidad de varianzas, para estar seguros de dicha conclusión se realizó una prueba de Levene.

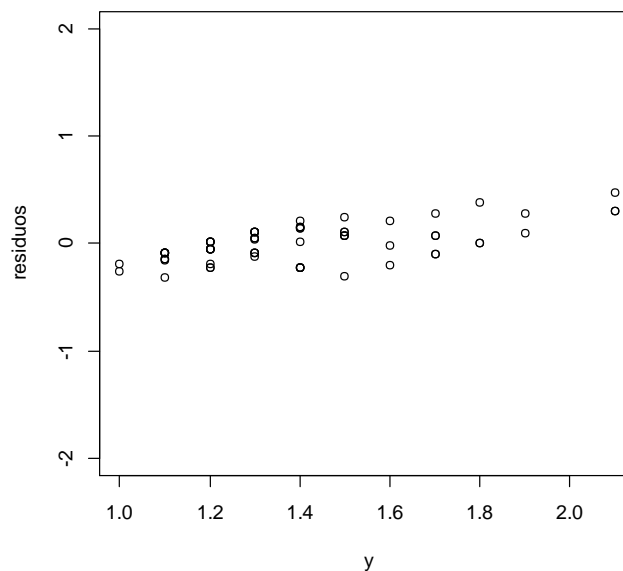
**Tabla 11.** Prueba de Levene.

	DF	PILLAI	APROXIMACIÓN F	NÚM F	DEN F	P VALOR
TRATAMIENTO	1	0,6189	4,8723	4	12	0,0144
RESIDUALES	15					

Como se observa el p-valor es inferior al nivel de significancia 0.05, es decir, que no hay igualdad de varianzas, ahora se observa si hay independencia entre los datos.

### 6.3.3. Prueba De Independencia

**Figura 9.** Prueba de independencia.



Fuente: R-Project (2017)

En la Figura anterior se observa que no hay independencia, además percibimos que los residuos poseen dependencia positiva, ya que existe correlación entre los datos. Por lo tanto, se concluye que el método de MANOVA no se ajusta a datos con medidas repetidas.

Ahora estudiaremos si los modelos mixtos pueden ser la herramienta útil al analizar este tipo de datos, ya que este método estadístico sirve para analizar datos correlacionados.

#### 6.4. MODELOS LINEALES MIXTOS

El análisis de los modelos mixtos está basado en deducir y estimar el modelo que, con un mínimo número de parámetros, se ajusta a los datos. Esto se logra modelando los datos con diferentes tipos de correlación, asumiendo varianzas homogéneas o heterogéneas y efectos aleatorios; seguido de estos se escoge el mejor modelo entre los propuestos, ya sea por el criterio el AIC o el BIC, depende del enfoque que se tenga del estudio de los datos.

Siguiendo con nuestro análisis, vamos a poner a prueba los siguientes modelos mixtos:

- a.** Errores independientes y homoscedásticos
- b.** Con semana como factor, estructura de simetría compuesta entre los errores del mismo animal y varianza residual constante en el tiempo
- c.** Sin semana como factor, estructura de simetría compuesta entre los errores del mismo animal y varianza residual constante en el tiempo
- d.** Sin estructura para las correlaciones entre errores provenientes del mismo animal y varianza residual constante en el tiempo.
- e.** Con semana como factor, estructura de correlación Autorregresiva de orden 1 entre los errores del mismo animal y varianza residual constante en el tiempo.
- f.** Sin semana como factor, estructura de correlación Autorregresiva de orden 1 entre los errores del mismo animal y varianza residual constante en el tiempo.
- g.** Con semana como factor, estructura de correlación Autorregresiva de orden 1 entre los errores del mismo animal y varianza residual diferente en los distintos tiempos
- h.** Con semana como factor, estructura de correlación Autorregresiva de orden 1 entre los errores del mismo animal y varianza diferente en los distintos tratamientos
- i.** Sin semana como factor, estructura de correlación Autorregresiva de orden 1 entre los errores del mismo animal y varianza residual diferente en los distintos tiempos

- j. Sin semana como factor, estructura de correlación Autorregresiva de orden 1 entre los errores del mismo animal y varianza diferente en los distintos tratamientos

A continuación, se muestra los resultados de los modelos a analizar.

**Tabla 12.** AIC y BIC de los modelos escogidos.

<b>Modelo</b>	<b>LOGLIK</b>	<b>AIC</b>	<b>BIC</b>
<b>MODELO 1</b>	9,58	-1,15	17,69
<b>MODELO 2</b>	25,14	-30,29	-9,34
<b>MODELO 3</b>	29,96	-47,93	-34,98
<b>MODELO 4</b>	33,2	-36,4	-4,99
<b>MODELO 5</b>	31,19	-42,37	-21,43
<b>MODELO 6</b>	36,39	-60,78	-47,83
<b>MODELO 7</b>	34,34	-42,68	-15,45
<b>MODELO 8</b>	33,78	-45,57	-22,53
<b>MODELO 9</b>	39,68	-60,36	-41,93
<b>MODELO 10</b>	37,88	-61,76	-46,65

Como se muestra en la tabla anterior hay dos modelos que pueden ser el indicado para nuestro estudio, ya que poseen el AIC y el BIC más bajo a comparación de los demás, y son los siguientes:

- a. Modelo 6.
- b. Modelo 10.

Para solucionar este problema se realiza una prueba de cociente de verosimilitud con los dos modelos anteriores, tomando en cuenta que se toma el modelo “1” como modelo reducido y el modelo “2” como modelo completo.

Formula:

$$\text{Cociente de verosimilitud} = 2 * [\text{LOGLIK (modelo completo)} - \text{LOGLIK (modelo reducido)}] \quad (83)$$

$$\begin{aligned}
&= 2*(37,88 - 36,39) \\
&= 2*(1,49) \\
&= 2,976
\end{aligned}$$

Para realizar esta prueba se debe tener en cuenta las siguientes hipótesis:

$H_0$ : Que el modelo reducido es mejor, modelo 6 (*Sin semana como factor*, estructura de correlación Autorregresiva de orden 1 entre los errores del mismo animal y varianza residual constante en el tiempo).

$H_1$ : Que el modelo completo es mejor, modelo 10 (*Sin semana como factor*, estructura de correlación Autorregresiva de orden 1 entre los errores del mismo animal y varianza diferente en los distintos tratamientos).

**Tabla 13.** Resultados del coeficiente de verosimilitud.

	MODELO	DF	AIC	BIC	LOGLIK	TEST	L.RADIO	P VALOR
<b>MODELO 6</b>	1	6	- 60,78	- 47,83	36,39			
<b>MODELO 10</b>	2	7	- 61,76	- 46,65	37,88	1 VS 2	2,9763	0,0845

De lo anterior, se concluye que el modelo reducido el modelo 6 es mejor a comparación del modelo completo el modelo 10, es decir, se acepta la hipótesis nula, ya que el p-valor es superior a 0.05. Por lo tanto, el modelo *Sin semana como factor*, estructura de correlación Autorregresiva de orden 1 entre los errores del mismo animal y varianza residual constante en el tiempo es el modelo que mejor se ajusta a estos datos. A continuación, se realiza toda la inferencia del modelo escogido.

El análisis de varianza del modelo escogido se muestra a continuación

**Tabla14.** ANOVA del modelo mixto 6.

	NUM F	F - VALOR	P VALOR	-
<b>INTERCEPTO</b>	1	1660,2351	0,0001	
<b>SEMANA</b>	1	73,0478	0,0001	
<b>TRATAMIENTO</b>	1	13,3711	0,0005	
<b>SEMANA*TRATAMIENTO</b>	1	14,807	0,0003	

Podemos concluir que si hay efecto significativo de interacción (valor  $p=0,0003$ ) entre los niveles de semana y tratamiento.

Ahora, se calculará la matriz de correlación a partir del  $\Phi = 0.7460$

$$AR(1) = \begin{bmatrix} 1 & 0.7460 & 0.5565 & 0.4151 \\ 0.4151 & 1 & 0.7460 & 0.5565 \\ 0.5565 & 0.4151 & 1 & 0.7460 \\ 0.7460 & 0.5565 & 0.4151 & 1 \end{bmatrix} \quad (84)$$

Ahora, vamos a realizar la prueba LSD Fisher para interacciones.

**Tabla 15.** Prueba LSD de Fisher por interacciones

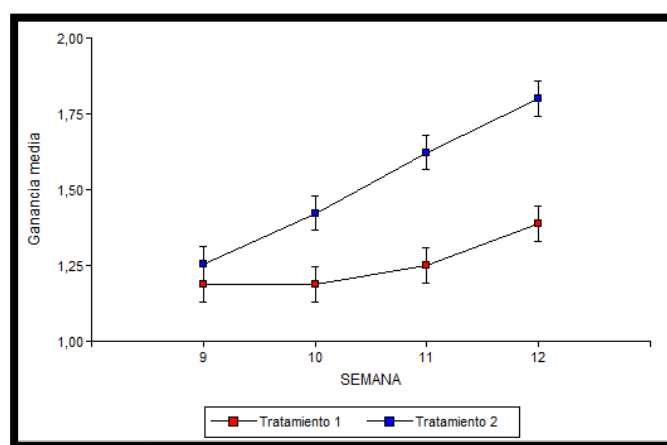
**LSD Fisher (Alfa=0,05)**

SEMANA	TRATAMIENTO	Medias	E.E.				
12	2	1,80	0,06	A			
11	2	1,62	0,06		B		
10	2	1,42	0,06			C	
12	1	1,39	0,06			C	D
9	2	1,26	0,06				D E
11	1	1,25	0,06				E
9	1	1,19	0,06				E
10	1	1,19	0,06				E

*Medias con una letra común no son significativamente diferentes ( $p > 0,05$ )*

Según la prueba LDS de Fisher se observa que la interacción (SEMANA 12\* TRATAMIENTO 2), es mejor comparándola con las demás interacciones, ya que fue la semana y tratamiento de mayor ganancia de peso registrado; además observemos que las interacciones (SEMANA10 \*TRATAMIENTO1) fue la que obtuvo menor ganancia de peso.

**Figura 10.** Figura LSD Fisher



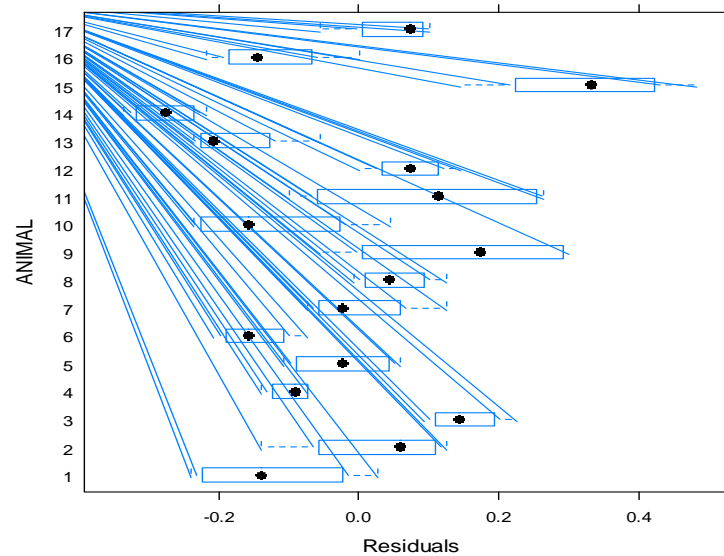
Fuente: R-Project (2017)

En la Figura podemos terminar de comprobar que la interacción es mejor; por ello, como conclusión se dice que la interacción (SEMANA 12 \* TRATAMIENTO 2), fue la que obtuvo mayor ganancia de peso.

Falta el gráfico de residuos del modelo escogido (modelo 6)

Por último, miremos la siguiente Figura,

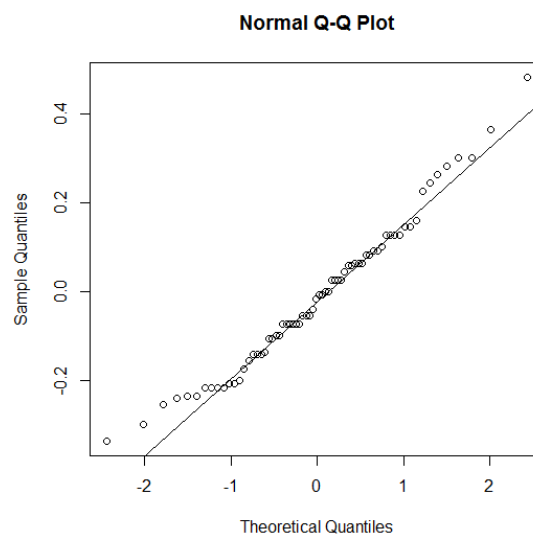
**Figura 11.** Variabilidad por animal.



Fuente: R-Project (2017)

Según lo observado en la Figura vemos que existe gran variabilidad entre los animales. Ahora, miremos cómo se comporta la normalidad y homogeneidad con el modelo escogido, pero primero se calcula los residuos y los predichos para realizar dichos análisis, después de realizar los cálculos se realizaron la Figura de normalidad homogeneidad,

**Figura 12.** Figura de normalidad

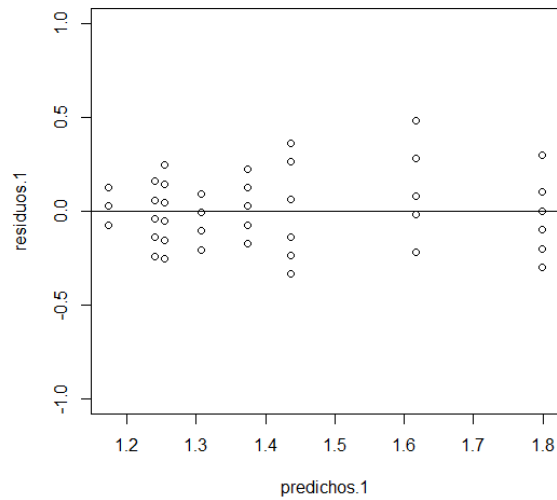


Fuente: R-Project (2017)



Como se refleja en la Figura anterior hay normalidad en el modelo escogido.

**Figura 13.** Prueba de Homoscedasticidad.



Fuente: R-Project (2017)

Como se observa en la Figura anterior hay Homoscedasticidad de varianzas en el modelo escogido.

## 7. CONCLUSIONES

Ya habiendo culminado el análisis de datos de medidas repetidas con metodologías estadísticas como el análisis de varianza (ANOVA), análisis multivariado (MANOVA) y los modelos lineales mixto (MLM), es viable concluir lo siguiente:

- Se efectuó un minucioso análisis de datos utilizando el software estadístico R-Project, logrando probar que procedimiento estadístico bajo la teoría de los modelos mixtos puede ser más útil y viable enfrentando datos con correlación como las medidas repetidas.
- Los datos con medidas repetidas utilizados en las aplicaciones son datos que están correlacionados entre sí, demostrado en la Figura 2, donde se expone la variabilidad de cada animal, además en la Figura 1 se puede ver que el tratamiento 2 tuvo mayor ganancia de peso a comparación del tratamiento 1.
- El procedimiento estadístico de análisis de varianza (ANOVA), no es el adecuado para modelar este tipo de datos, ya que se incumple el supuesto de independencia, pero a pesar de lo anterior si se cumplen los supuestos de normalidad y Homoscedasticidad, no obstante sino se cumple el supuesto de independencia el análisis de varianza no puede ser una herramienta útil en el estudio de los datos, esto ocurre ya que los datos están correlacionados.
- Al igual que en el análisis de varianza, el método estadístico de análisis multivariado (MANOVA), pasa lo mismo, se cumplen los supuestos de normalidad y Homoscedasticidad, pero no se cumple el supuesto de independencia, a causa de que los datos de medidas repetidas son datos correlacionados entre sí, por tal motivo el análisis multivariado (MANOVA), no es una herramienta útil a la hora de analizar este tipo de datos.

- Se utilizó los modelos lineales mixtos, a razón de que este tipo de metodología estadística, modela datos con diferentes tipos de correlación y varianzas heterogéneas u homogéneas; se llegó a que los datos que fueron de nuestro estudio poseen correlación auto regresiva de orden 1, se afirma lo anteriormente dicho porque se trabajó con diversos tipos de correlación y teniendo en cuenta el criterio de información de Akaike (AIC), este tipo de correlación se ajusta al tipo de datos que se utilizó en nuestro análisis de medidas repetidas.
- Se realizó una prueba LSD Fisher por semana, tratamiento y su interacción (semana\*tratamiento), obteniendo como resultado que la semana 12 fue la mejor semana de ganancia de peso comparándola con las semanas 9,10 y 11, además se obtuvo que el tratamiento 2 fue el que mejor y mayor ganancia de peso tuvo y por último se llegó a que la interacción (semana 12 \* tratamiento 2) fue la que más obtuvo mayor ganancia de peso a comparación de las demás interacciones.

## RECOMENDACIONES Y TRABAJOS FUTUROS

Una de las metodologías que ha estado tomando bastante fuerza en los últimos años es la de los datos funcionales (FDA), se propone entonces estudiar las medidas repetidas en el tiempo mediante esta técnica, tomando las mediciones del mismo sujeto como una función y ver si esta nueva metodología nos arroja un mejor resultado que la de los modelos lineales mixtos.

## REFERENCIAS BIBLIOGRAFICAS

- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. UR <http://www.R-project.org/>.
- Arнау-Gras, J. (2007). Estudios longitudinales de medidas repetidas. Modelos de diseño y de análisis. Avances en medición, 5, 9-26. España, Barcelona.
- Kaps, M., & Lamberson, W. (2009). Biostatistics for animal science. CABI Publishing.
- Carmona, F. (2005). Modelos lineales. Universidad de Barcelona, España, Barcelona.
- Cox, D. (1958). Planning of experiment. New York: John Wiley & Sons, Ltd.
- Díaz, L. (2002). Estadística multivariada: inferencia y métodos. Facultad de Ciencias, Universidad Nacional de Colombia. Bogotá, D.C.
- Di Rienzo, A., Macchiavelli, R., & Casanoves, F. (2009). Modelos mixtos en Infostat. Tutorial, Grupo InFoStat, FCA, Universidad Nacional de Córdoba, AR.
- García, P., Rojas, P., & Seco, G. (2007). Cómo elegir la mejor prueba estadística para realizar un diseño de medidas repetidas. Revista Internacional de Psicología Clínica y de Salud, 7(1), 153-175. España & Chile.
- Gómez, S., Torres, V., García, Y., & Navarro, J. (2012). Procedimientos estadísticos más utilizados en el análisis de medidas repetidas en el tiempo en el sector agropecuario. Revista Cubana de Ciencia Agrícola, 46(1), 1.
- Falcó, M. (2009). Herramientas estadísticas-comparación de más de dos muestras: ANOVA (parte I). Revista Comillas. Modulo XIII. Madrid.
- Kaps, M., & Lamberson, W. (2009). Biostatistics for animal science, 2 ed. CABI.
- Littell, R., Stroup, W., Milliken, G., Wolfinger, R., & Schabenberger, O. (2006). SAS for mixed models. Capitulo V. SAS institute. USA.
- Llobell, J., Navarro, M., & Pérez, J. (1996). Manual de psicología experimental, Capitulo III: Diseño de medidas repetidas. Ariel Psicología.
- Maxwell, S., & Delaney, H. (2004). Designing experiments and analyzing data: A model comparison perspective (Vol. 1). Psychology Press. London, New York.
- Mena, M. (2004). Alternativas de análisis estadístico en los diseños de medidas repetidas. Psicothema, 16(3), 509-518. España.

- Merino, A., & Castellanos, R. S. M. (1998). Análisis de datos en psicología II. 2 ed. Edit. Pirámide. España.
- Pulido, H., De la Vara, R., González, P., Martínez, C., & Pérez, M. (2012). Análisis y diseño de experimentos. McGraw-Hill Interamericana.
- Ruiz, M. (2004) Análisis de Medidas Repetidas. Revista Medicina Clínica 4; 122(Suplemento 1):51-8 Barcelona; España.
- Seoane, J. (2004). ¿Modelos mixtos (lineales)? Una introducción para el usuario temeroso. Etologuía. Ed 24. Madrid, España.
- Steel, R., Torrie, J., & Martínez, R. (1985). Bioestadística: principios y procedimientos (Vol. 2). McGraw-Hill Interamericana S.A. Bogotá.

# ANEXOS

## **Anexo A.** Rutinas hechas en el Software R-Project

```
datos =read.delim("clipboard")
datos
str(datos)
attach(datos)
interaction.plot(SEMANA,TRATAMIENTO,GANANCIA,
col=1:2,lwd=3,type="b",pch=1,ylab="GANANCIA MEDIA",ylim=c(1.0,2.0),
,xlab="SEMANA",main="Tendencia de la ganacia de los animales por tratamiento")
interaction.plot(SEMANA,ANIMAL,GANANCIA,col=1:17,
lwd=1,type="l",ylab="GANANCIA MEDIA",lty=1,
,xlab="SEMANA",main="Tendencia de la ganacia por animal animales")
```

### **##### ANALISIS UNIVARIADO (ANOVA)**

```
SEMANA=factor(SEMANA)
TRATAMIENTO=factor(TRATAMIENTO)
ANIMAL=factor(ANIMAL)
anova=aov(GANANCIA~TRATAMIENTO+SEMANA+TRATAMIENTO:SEMANA)
summary(anova)
residuos=residuals(anova)
residuos
predichos=fitted(anova)
predichos
```

### **PRUEBA DE NORMALIDAD**

```
shapiro.test(residuos)
qqnorm(residuos)
qqline(residuos)
```

### **PRUEBA DE HOMOCEDASTICIDAD**

```
plot(predichos,residuos,ylim=c(-2,2))
abline(h=0)
abs.res=abs(residuos)
levene=aov(abs.res~TRATAMIENTO+SEMANA+TRATAMIENTO:SEMANA)
summary(levene)
```



## **PRUEBA DE INDEPENDENCIA**

```
plot(residuos,SEMANA,ylim=c(-1,6))
```

## **##### ANALISIS MULTIVARIADO (MANOVA)**

```
y=cbind(SEMANA.9,SEMANA10,SEMANA.11,SEMANA.12)
```

```
TRATAMIENTO= as.factor(TRATAMIENTO)
```

```
TRATAMIENTO
```

```
M1=manova(y~TRATAMIENTO)
```

```
M1
```

```
summary(M1,test="Wilks")
```

```
residuos=residuals(M1)
```

```
residuos
```

```
predichos=fitted(M1)
```

```
predichos
```

## **PRUEBA DE NORMALIDAD**

```
shapiro.test(residuos)
```

```
qqnorm(residuos)
```

```
qqline(residuos)
```

## **PRUEBA DE HOMOCEDASTICIDAD**

```
plot(predichos,residuos,ylim=c(-1,1))
```

```
abline(h=0)
```

```
abs.res=abs(residuos)
```

```
levene=M1
```

```
summary(levene)
```

## **PRUEBA DE INDEPENDENCIA**

```
plot(residuos,SEMANA,ylim=c(0,3))
```

## **##### MODELOS LINEALES MIXTOS**

**##### Modelo1: Varianzas Residuales Homogéneas y errores independientes.**

```
SEMANA=factor(SEMANA)
```

```
TRATAMIENTO=factor(TRATAMIENTO)
```

```
ANIMAL=factor(ANIMAL)
```

```
library(nlme)
```

```

modelo1=glS(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO)
summary(modelo1)
anova(modelo1)

```

**##### Modelo2: Varianzas Residuales Homogéneas con semana como factor y estructura de correlación Simetría Compuesta.**

```

SEMANA=as.factor(SEMANA)
TRATAMIENTO=as.factor(TRATAMIENTO)
ANIMAL=as.factor(ANIMAL)
modelo2=glS(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO
,correlation=corCompSymm(form=~1|ANIMAL))
summary(modelo2)
anova(modelo2)

```

**##### Modelo3: Varianzas Residuales Homogéneas sin semana como factor y estructura de correlación Simetría Compuesta.**

```

TRATAMIENTO=as.factor(TRATAMIENTO)
ANIMAL=as.factor(ANIMAL)
modelo3=glS(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO
,correlation=corCompSymm(form=~1|ANIMAL))
summary(modelo3)
anova(modelo3)

```

**##### Modelo4: Varianzas Residuales Homogéneas sin estructura.**

```

modelo4=glS(GANANCIA~1+TRATAMIENTO+SEMANA+SEMANA:TRATAMIENTO
,correlation=corSymm(form=~as.integer(rank(as.numeric(as.character(SEMANA))))|ANIMAL))
summary(modelo4)
anova(modelo4)

```

**##### Modelo5: Varianzas Residuales Homogéneas con semana como factor y estructura de correlación Auto correlación de Orden 1.**

```

SEMANA=factor(SEMANA)
TRATAMIENTO=factor(TRATAMIENTO)
ANIMAL=factor(ANIMAL)

```

```
library(nlme)
modelo.5=glS(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO
,method="REML",correlation=corAR1(form=~as.integer(as.character(SEMANA))|ANIMA
L))
summary(modelo.5)
anova(modelo.5)
```

**##### Modelo6: Varianzas Residuales Homogéneas sin semana como factor y estructura de correlación Auto correlación de Orden 1.**

```
TRATAMIENTO=factor(TRATAMIENTO)
ANIMAL=factor(ANIMAL)
library(nlme)
modelo.6=glS(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO
,method="REML",correlation=corAR1(form=~as.integer(as.character(SEMANA))|ANIMA
L))
summary(modelo.6)
anova(modelo.6)
```

**#### Se observara si es posible de que haya varianzas heterogéneas.**

```
#box- plots of residuales por semana
plot(modelo1, SEMANA ~ resid(.),ylim=c(0,5),main=' plots of residuales por SEMANA')
#box- plots of residuales por tratamiento
plot(modelo1, TRATAMIENTO ~ resid(.),ylim=c(0,3),main=' plots of residuales por
TRATAMIENTO')
```

**#### MODELOS HETEROGENEOS**

**### Modelo7. Varianzas Residuales Homogéneas con semana como factor, heterogeneidad para semana y estructura de correlación Auto correlación de Orden 1.**

```
SEMANA=factor(SEMANA)
TRATAMIENTO=factor(TRATAMIENTO)
ANIMAL=factor(ANIMAL)
library(nlme)
```

```

modelo.7=glS(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO,
weights=varComb(varIdent(form=~1|SEMANA)),method="REML"
,correlation=corAR1(form=~as.integer(as.character(SEMANA))|ANIMAL))
summary(modelo.7)
anova(modelo.7)

```

**### Modelo8. Varianzas Residuales Homogéneas con semana como factor, heterogeneidad para tratamiento y estructura de correlación Auto correlación de Orden 1.**

```

SEMANA=factor(SEMANA)
TRATAMIENTO=factor(TRATAMIENTO)
ANIMAL=factor(ANIMAL)
library(nlme)
modelo.8=glS(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO,
weights=varComb(varIdent(form=~1|TRATAMIENTO)),method="REML"
,correlation=corAR1(form=~as.integer(as.character(SEMANA))|ANIMAL))
summary(modelo.8)
anova(modelo.8)

```

**#### Modelo9. Varianzas Residuales Homogéneas sin semana como factor heterogeneidad para semana y estructura de correlación Auto correlación de Orden 1.**

```

TRATAMIENTO=factor(TRATAMIENTO)
ANIMAL=factor(ANIMAL)
library(nlme)
modelo.9=glS(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO,
weights=varComb(varIdent(form=~1|SEMANA)),method="REML"
,correlation=corAR1(form=~as.integer(as.character(SEMANA))|ANIMAL))
summary(modelo.9)
anova(modelo.9)

```

**#### Modelo10. Varianzas Residuales Homogéneas sin semana como factor, heterogeneidad para tratamiento y estructura de correlación Auto correlación de Orden 1.**

```

TRATAMIENTO=factor(TRATAMIENTO)
ANIMAL=factor(ANIMAL)
library(nlme)
modelo.10=glS(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO,
weights=varComb(varIdent(form=~1|TRATAMIENTO)),method="REML"
,correlation=corAR1(form=~as.integer(as.character(SEMANA))|ANIMAL))
summary(modelo.10)
anova(modelo.10)

```

#### ##### **COCIENTE DE VEROSIMILITUD**

```
anova(modelo.6, modelo.10)
```

#### ##### **Análisis Del Modelo6: Varianzas Residuales Homogéneas sin semana como factor y estructura de correlación Auto correlación de Orden 1.**

modelo6: Varianzas Residuales Homogéneas sin semana como factor y estructura de correlación Auto correlación de Orden 1.

```

TRATAMIENTO=factor(TRATAMIENTO)
ANIMAL=factor(ANIMAL)
library(nlme)
modelo.6=glS(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO
,method="REML",correlation=corAR1(form=~as.integer(as.character(SEMANA))|ANIMA
L))
summary(modelo.6)
anova(modelo.6)

```

#### ##### **TEST DE LSD DE FISHER**

```

library(agricolae)
Y=LSD.test(anova,"SEMANA")
Z=LSD.test(anova,"TRATAMIENTO")

```

#### ### **GRAFICA PARA ANALIZAR LA VARIABILIDAD.**

```
plot(modelo.6, ANIMAL ~ resid(.))
```

#### ### **GRAFICAS DE NORMALIDAD Y HOMOGENEIDAD.**

```
TRATAMIENTO=factor(TRATAMIENTO)
```

```
ANIMAL=factor(ANIMAL)
```

```
library(nlme)
```

```
modelo.6=glms(GANANCIA~1+SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO  
,method="REML",correlation=corAR1(form=~as.integer(as.character(SEMANA))|A  
NIMAL))
```

```
summary(modelo.6)
```

```
anova(modelo.6)
```

```
residuos.1=residuals(modelo.6)
```

```
residuos.1
```

```
predichos.1=fitted(modelo.6)
```

```
predichos.1
```

#### **PRUEBA DE NORMALIDAD**

```
shapiro.test(residuos.1)
```

```
qqnorm(residuos.1)
```

```
qqline(residuos.1)
```

#### **PRUEBA DE HOMOCEDASTICIDAD**

```
plot(predichos.1,residuos.1,ylim=c(-1,1))
```

```
abline(h=0)
```

```
abs.res.1=abs(residuos.1)
```

```
levene=aov(abs.res.1~SEMANA+TRATAMIENTO+SEMANA:TRATAMIENTO)
```

	<b>SISTEMA DE GESTION DE LA CALIDAD</b>  <b>FORMATO DE AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL</b>	Página 1 de 3
		Código: GB-P04-F03
		Versión: 02

Los suscritos:

Andrés Mauricio Bernal Pérez	con C.C N°	1110516832
Jhon Alexander rozo Forero	con C.C N°	1110435670
	con C.C N°	
	con C.C N°	
	con C.C N°	

Manifiesto (an) la voluntad de:

Autorizar

☒

No Autorizar

☐

Motivo:

La consulta en físico y la virtualización de **mi OBRA**, con el fin de incluirlo en el repositorio institucional de la Universidad del Tolima. Esta autorización se hace sin ánimo de lucro, con fines académicos y no implica una cesión de derechos patrimoniales de autor.

Manifestamos que se trata de una OBRA original y como de la autoría de LA OBRA y en relación a la misma, declara que la UNIVERSIDAD DEL TOLIMA, se encuentra, en todo caso, libre de todo tipo de responsabilidad, sea civil, administrativa o penal (incluido el reclamo por plagio).

Por su parte la UNIVERSIDAD DEL TOLIMA se compromete a imponer las medidas necesarias que garanticen la conservación y custodia de la obra tanto en espacios físico como virtual, ajustándose para dicho fin a las normas fijadas en el Reglamento de Propiedad Intelectual de la Universidad, en la Ley 23 de 1982 y demás normas concordantes.

La publicación de:

Trabajo de grado	<input checked="" type="checkbox"/>	Artículo	<input type="checkbox"/>	Proyecto de Investigación	<input type="checkbox"/>
Libro	<input type="checkbox"/>	Parte de libro	<input type="checkbox"/>	Documento de conferencia	<input type="checkbox"/>
Patente	<input type="checkbox"/>	Informe técnico	<input type="checkbox"/>		
Otro: (fotografía, mapa, radiografía, película, video, entre otros)					<input type="checkbox"/>

Fecha Versión 02: 04-11-2016

	<b>SISTEMA DE GESTION DE LA CALIDAD</b>  <b>FORMATO DE AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL</b>	Página 2 de 3
		Código: GB-P04-F03
		Versión: 02

Producto de la actividad académica/científica/cultural en la Universidad del Tolima, para que, con fines académicos e investigativos, muestre al mundo la producción intelectual de la Universidad del Tolima. Con todo, en mi condición de autor me reservo los derechos morales de la obra antes citada con arreglo al artículo 30 de la Ley 23 de 1982. En concordancia suscribo este documento en el momento mismo que hago entrega del trabajo final a la Biblioteca Rafael Parga Cortes de la Universidad del Tolima.

De conformidad con lo establecido en la Ley 23 de 1982 en los artículos 30 “**...Derechos Morales. El autor tendrá sobre su obra un derecho perpetuo, inalienable e irrenunciable**” y 37 “**...Es lícita la reproducción por cualquier medio, de una obra literaria o científica, ordenada u obtenida por el interesado en un solo ejemplar para su uso privado y sin fines de lucro**”. El artículo 11 de la Decisión Andina 351 de 1993, “**los derechos morales sobre el trabajo son propiedad de los autores**” y en su artículo 61 de la Constitución Política de Colombia.

- Identificación del documento:
  - **VENTAJAS Y DESVENTAJAS DEL ANÁLISIS DE UN DISEÑO DE MEDIDAS REPETIDAS MEDIANTE ANÁLISIS DE VARIANZA UNIVARIADO, MULTIVARIADO Y MODELOS LINEALES MIXTOS**
- Trabajo de grado como requisito parcial para optar al título de:
 

**Profesional en Matemáticas con Énfasis en Estadística**

- 
- Proyecto de Investigación correspondiente al Programa (No diligenciar si es opción de grado “Trabajo de Grado”):

---

- Informe Técnico correspondiente al Programa (No diligenciar si es opción de grado “Trabajo de Grado”):

---

- Artículo publicado en revista:

---

- Capítulo publicado en libro:


---

- Conferencia a la que se presentó:

---

Fecha Versión 02: 04-11-2016




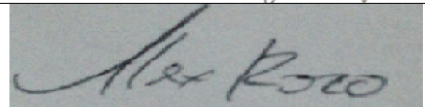
	<b>SISTEMA DE GESTION DE LA CALIDAD</b>  <b>FORMATO DE AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL</b>	Página <b>3</b> de <b>3</b>
		Código: GB-P04-F03
		Versión: 02

Quienes a continuación autentican con su firma la autorización para la digitalización e inclusión en el repositorio digital de la Universidad del Tolima, el:

Día: **22** Mes: **Agosto** Año: **2017**

Autores:

Firma

Nombre:	<b>Andrés Mauricio Bernal Pérez</b>		C.C.	<b>1110516832</b>
Nombre:	<b>Jhon Alexander Roza Forero</b>		C.C.	<b>1110495670</b>
Nombre:			C.C.	
Nombre:			C.C.	

El autor y/o autores certifican que conocen las derivadas jurídicas que se generan en aplicación de los principios del derecho de autor.

Fecha Versión 02: 04-11-2016